DOCUMENT RESUME

ED 457 425 CG 031 161

AUTHOR	Walz, Garry R., Ed.; Bleuer, Jeanne C., Ed.
TITLE	Assessment: Issues and Challenges for the Millennium.
INSTITUTION	ERIC Clearinghouse on Counseling and Student Services, Greensboro, NC.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
ISBN	ISBN-1-56109-091-3
PUB DATE	2001-00-00
NOTE	392p.
CONTRACT	ED-99-CO-0014
AVAILABLE FROM	CAPS Publications, School of Education, University of North Carolina at Greensboro, 201 Ferguson Building, Greensboro, NC 27402-6171 (Item no. EC256, $29.95) Tel: 336-334-4114; Tel: 800-414-9769 (Toll Free); Fax: 336-334-4116; Web site: http://ericcass.uncg.edu; e-mail: ericcass@uncg.edu.
PUB TYPE	Books (010) -- Collected Works - General (020) -- ERIC Publications (071)
EDRS PRICE	MF01/PC16 Plus Postage.
DESCRIPTORS	Counseling; *Counselor Role; *Counselors; *Educational Assessment; Educational Theories; Elementary Secondary Education; *Evaluation; Higher Education; History; *Measures (Individuals); *Testing; Theory Practice Relationship

ABSTRACT

The collection of papers presented in this monograph provides an update on developments in assessment for counselors, practitioners, and researchers in all types of educational settings. The papers are grouped into five sections. Section 1, on emerging issues, includes topics on testing theories presently used, and some thoughts for the future. Section 2 concerns issues of counselor education in higher education. Section 3 discusses educational assessment in elementary and secondary schools, and the role of the counselor in school guidance. Section 4 considers testing modifications and accommodations that are in practice with the special needs population. Section 5 reviews special assessment topics. Examples of these topics include the rights of the test takers; communicating test results; writing multiple choice tests; and use of specific types of tests. Also included are reprints of 33 ERIC/CASS Digests on Assessment in Counseling and Therapy produced for the Assessment '95 ERIC/CASS and the Association for Assessment in Counseling national conference. With the publication of this monograph, it is hoped that a wide range of practitioners and researchers will have access to current information that is not readily available to them. (Contains over 300 references.) (JDM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 457 425

ASSESSMENT

Issues and Challenges for the Millennium

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CG031161

BEST COPY AVAILABLE

Editors:
Garry R. Walz, Ph.D. & Jeanne C. Bleuer, Ph.D.

Assessment

Issues and Challenges for the Millennium

Editors
Garry R. Walz, Ph.D., NCC
Jeanne C. Bleuer, Ph.D., NCC

Copyright © 2001 CAPS Publications
PO Box 35077
Greensboro, NC 27425-5077

All rights reserved

ISBN 1-56109-091-3

This publication was funded in part by the U.S. Department of Education, Office of Educational Research and Improvement, Contract No. ED-99-CO-0014. Opinions expressed in this publication do not necessarily reflect the positions of the U.S. Department of Education, OERI, or ERIC/CASS.

Assessment
Issues and Challenges for the Millennium

Preface .... *vii*

Emerging Issues

Chapter One *3*
Contemporary Assessment: A Hotbed of Issues and Challenges
Edwin L. Herr

Chapter Two *29*
An Emerging Paradigm of Testing
Lorin Letendre

Chapter Three *43*
Responding to Testing Needs in the Twenty-First Century With an Old Tool
Lawrence M. Rudner

Counselor Education

Chapter Four *55*
Assessment in Counselor Education: Admissions, Retention, and Capstone Experiences
Irene M. Ametrano and Sue A. Stickel

Chapter Five *65*
Revitalizing the Assessment Course in the Counseling Curriculum
Albert B. Hood

Chapter Six *77*
The Pedagogical Basis for Multifaceted Assessment in Counselor Education
Barbara D. Yunker and Mary E. Stinson

K–12 Assessment and Counseling

Chapter Seven *87*
Addressing Fragmentation: Building Integrated Services for Student Support
Jackie M. Allen

Chapter Eight *95*
Competencies in Assessment and Evaluation for School Counselors
Joint Committee of the American School Counselor Association (ASCA) and the Association for Assessment in Counseling (AAC), chaired by Patricia Elmore

Chapter Nine *101*
Assessing the Effectiveness of School Guidance Programs: Program, Personnel, and Results Evaluation
Norman C. Gysbers

Special Populations

Chapter Ten *115*
Assessing Diverse Populations
Courtland C. Lee

Chapter Eleven *125*
Equity Issues in the Assessment of Individuals With Visual or Hearing Impairments
Ruth B. Ekstrom

Chapter Twelve *131*
Modifying Tests for Students With Disabilities
Douglas K. Smith

Special Assessment Topics

Chapter Thirteen *143*
Development of a Statement of Test Takers' Rights and Responsibilities
Kurt F. Geisinger

Chapter Fourteen *163*
Counseling Outcome Research: Making Practical Choices for Real-World Applications
Darcy H. Granello and Paul F. Granello

Chapter Fifteen *173*
Communicating Assessment Results in the Counseling Interview
Albert B. Hood

Chapter Sixteen *185*
Response to Surveys by Electronic and Traditional Means
David J. Lundberg

Chapter Seventeen *189*
Using the Internet to Enhance Test Interpretation
James P. Sampson Jr.

Chapter Eighteen *203*
Use of the Kaufman Adolescent and Adult Intelligence Test (KAIT) in the New Millennium
Douglas K. Smith

Chapter Nineteen *215*
Writing Multiple-Choice Test Items
Nicholas A. Vacc, Larry C. Loesch, and Ruth E. Lubik

Reprint of the Assessment '95 Conference Papers
Assessment in Counseling and Therapy Special Digest Collection
William D. Schafer, Guest Editor

Overview of Assessment in Counseling and Therapy *229*
William D. Schafer

Assessment Skills of Counselors, Principals, and Teachers *235*
James C. Impara

Assessment Skills for School Counselors *241*
William D. Schafer

Mental Health Counseling Assessment: Broadening One's Understanding of the Client and the Client's Presenting Concerns ... *247*
Gerald A. Juhnke

CACREP Accreditation: Assessment and Evaluation in the Standards and Process ... *253*
Carol L. Bobby and Joseph R. Kandor

The Role of Assessment in Counselor Certification ... *259*
Thomas Clawson

Assessment of Counselor Performance ... *265*
Larry C. Loesch

Evaluating School Guidance Programs ... *271*
Norman C. Gysbers

New Assessment Methods for School Counselors ... *277*
W. James Popham

Sound Performance Assessments in the Guidance Context ... *281*
Richard J. Stiggins

Portfolios for Assessment and Instruction ... *287*
Judith A. Arter, Vicki Spandel, and Ruth Culham

Emerging Student Assessment Systems for School Reform ... *293*
Edward Roeber

Assessment of Abilities ... *299*
Thomas F. Harrington

Interest Assessment ... *305*
Jo-Ida C. Hansen

Assessment of Self-Concept ... *309*
William Strein

Assessment of Temperament ... *313*
Hedwig Teglasi

Assessment of Preschool Children *319*
Nicholas A. Vacc and Sandra H. Ritter

Screening for Special Diagnoses *325*
Susan De La Paz and Steve Graham

Assessment in Career Counseling *329*
Dale J. Prediger

Assessing Career Certainty and Choice Status *335*
Paul J. Hartung

Gender Differences in Adolescent Career Exploration *339*
Helen S. Farmer

Assessing Employability Skills *345*
Thomas H. Saterfiel and Joyce R. McLarty

Assessing Career Development With Portfolios *349*
Juliette N. Lester and Nancy S. Perry

Ethics in Assessment *355*
Cynthia B. Schmeiser

Multicultural Assessment *361*
William E. Sedlacek and Sue H. Kim

Fairness in Performance Assessment *365*
Tony C. M. Lam

Computer-Assisted Testing in Counseling and Therapy *371*
James P. Sampson Jr.

Testing Students With Disabilities *375*
Kurt F. Geisinger and Janet F. Carlson

The School Psychologist's Role in School Assessment *381*
Sylvia Rosenfield and Deborah Nelson

Cooperation Between School Psychologists and Counselors in Assessment *387*
Douglas K. Smith

Using Buros Institute of Mental Measurements Materials in Counseling and Therapy *393*
Barbara S. Plake and Jane C. Conoley

Internet Resources for Guidance Personnel *399*
Liselle Drake and Lawrence M. Rudner

Locating and Evaluating Career Assessment Instruments *405*
Jerome T. Kapes

Inappropriate Statistical Practices in Counseling Research: Three Pointers for Readers of Research Literature *411*
Bruce Thompson

Tables

2.1. Comparison of Emergent and Established Matrices
3.1. Sample Probabilities of Correct and Incorrect Responses by Masters and Non-masters
3.2. Calculations for Probability of Mastery Based on Four Sample Responses
18.1. Subtests of the KAIT
18.2. Fluid Subtests on Various Cognitive Batteries
18.3. Effective Range of Standard Scores

Preface

This monograph offers an intriguing collection of manuscripts. They are attractive in terms of both the breadth of topics covered and the concepts presented. They make excellent reading and provide an update on new developments in assessment as well as compelling discussions of important topics. As even a quick review of the titles will readily suggest, the comprehensiveness of the papers is impressive. Scholars/researchers, counselor educators, practitioners, and graduate students will all find much to their liking and interest.

The papers have been grouped into five sections: (a) Emerging Issues, (b) Counselor Education, (c) K–12 Guidance and Counseling, (d) Special Populations, and (e) Special Topics. A unique feature of this monograph is the inclusion of 33 ERIC/CASS Digests on Assessment in Counseling and Therapy that were produced for the Assessment '95 national conference, offered jointly by ERIC/CASS and the Association for Assessment in Counseling (AAC). This collection of digests, edited by Dr. William Schafer and authored by an impressive group of assessment experts, provides succinct and cogent statements regarding a wide range of topics. Due to their limited distribution but still highly relevant content, we believe that they offer useful additional resources of value to both researchers and practitioners.

We are pleased to offer this monograph as continuing evidence of our interest at ERIC/CASS in sharing information on assessment as it relates to counseling, therapy, and human services. We believe that it will be a valuable addition to any collection of resources on assessment.

Garry R. Walz, Ph.D., NCC
Co-Director, ERIC/CASS
Professor Emeritus, University of Michigan

Jeanne C. Bleuer, Ph.D., NCC
Co-Director, ERIC/CASS

Emerging Issues

Contemporary Assessment: A Hotbed of Issues and Challenges
Edwin L. Herr

An Emerging Paradigm of Testing
Lorin Letendre

Responding to Testing Needs in the Twenty-First Century With an Old Tool
Lawrence M. Rudner

Chapter One

Contemporary Assessment: A Hotbed of Issues and Challenges

Edwin L. Herr[1]

Abstract

At all levels assessment has become a sociopolitical instrument, from school accountability assessments to worker evaluations, to program accountability and certification measures. The application of assessment to new purposes with potentially significant impact on individuals and groups raises old and new questions about how assessment should be conducted, by whom, and in what contexts. Concerns about potential ethnic or gender bias, the validity of inferences drawn from tests, the interaction between counseling and assessment, the professional ownership of assessment, computer applications in assessment and counseling, and related issues are reviewed.

The phrase "hotbed of issues and challenges" is certainly a concept that is apropos when looking at the status of assessment in terms of federal policy, the social or economic climate related to education, the role of testing in counseling, assessment as an intervention in its own right, or assessment as it is affected by technology. Such emphases, although not exhaustive, are important parts of the context in which the issues and challenges related to assessment can be framed at the beginning of the twenty-first century and into the next several decades. Given the lack of certainty we have about projections related to assessment in the years ahead, we may find it useful to raise a series of questions about assessment that will help us, as professionals, to engage in long-term strategic planning as a way of reducing the uncertainty we experience as we consider the likely changes ahead.

I am writing not as an assessment expert, which I am not, but rather as a long-time educational administrator and counselor who has been and is concerned about the results of assessment and the issues that affect such processes. Given these preliminary perspectives, let me try to construct an "external" view of assessment and some of the issues and challenges that I see embedded in such a view. First, I would argue that at present in the United States, assessment is being expected to perform roles and functions that are unprecedented in its history. As Terwilliger (1997) has observed:

> It is obvious to even the most casual reader of the literature on educational assessment that the field is currently undergoing a fundamental and profound transformation. The traditional concepts and methodologies associated with assessment are being questioned by a variety of critics including school reform advocates, subject matter experts, cognitive theorists, and others. In general, advocates for change recommend that assessments of achievement should be designed to reflect more precisely complex 'real-life' performances and problems than is possible with short-answer and choice-response questions that characterize many teacher-made tests. (p. 24)

Whether justified or not, implications for assessment pervade the language of federal and state policy directed to the healthcare industry, to manufacturing and financial processes, to both basic and higher education, and to the allocation of resources. Whether we are in a university or a school district, in community mental health services, or in other social and economic institutions, terms such as *data-driven, standards, performance indicators, continuous quality improvement, total quality management, accountability, benchmarking, strategic initiatives and strategic actions, competency, certification, accreditation,* and *licensure* have become standard vocabulary and operating processes that define much of our professional existence. Each of these processes, as they are implemented, embodies some form of assessment, measurement, or testing. This reality has promoted one anonymous wag to suggest that in the Constitution of the United States we have replaced the creed of the founding fathers that "We hold these truths to be self-evident" with the words "We hold these truths to be statistical."

Assessment in the United States is still young in chronological terms; depending on your assumptions, it dates, in operational terms, only to the first two decades of the twentieth century when Binet brought his intelligence test from France to the United States, or the Army Alpha tests were developed for use in World War I, or the original vocational interest measure was given preliminary attention at the first applied psychology department in the nation at what was then called

Carnegie Tech in Pittsburgh in 1915. Assessment in the United States has achieved much and grown as a science during the last seven or eight decades, but throughout the twentieth century, it has tended to be used for rather specific purposes and within restricted contexts. Even though there are exceptions to this rule, assessment has not been consistently defined by government policy as a sociopolitical instrument of national importance. The National Defense Education Act, the armed forces, and the federal enabling legislation for employment counselors and rehabilitation counselors have certainly emphasized testing—for example, to identify gifted adolescents who should be encouraged to enter science and mathematics in higher education, or to identify the performance capabilities of inductees into the military, the unemployed, or people with disabilities for whom specific training should be provided. In general, assessment *purposes* and uses have evolved incrementally as assessment knowledge and techniques have evolved, their purposes and uses in most cases have been limited, and they have not been the focus of federal policy debate.

At the moment, however, we appear to have developed a national love-hate climate surrounding assessment as such processes have become partisan political grist for public policy debates between political parties and the president of the United States or between political parties and special-interest groups. Examples of these issues were reflected in the headlines about President Clinton's commitments to national academic standards in reading and mathematics and voluntary national testing to determine whether these standards are being met, as opposed to the stance of Representative William Goodling, retiring chair of the House Education and Workplace Committee, who thinks federal money should be focused not on testing, but on better education (Hoff, 1997). In addition, Representative Goodling's plan would require the National Academy of Sciences to review all existing commercial tests to determine whether they can create "an equivalency scale to compare students' scores on them." Paralleling Representative Goodling's original request, President Clinton assigned the Governing Board of the National Assessment of Educational Progress to study his proposal for new national tests, their use, and their design (Lawton, 1997b). Although there have been recent compromises between the parties on the president's initiatives related to voluntary testing, other federal and state legislation—including the goals of President George W. Bush for educational reform and accountability—continue to elevate assessment into a major strategy. For example, as of the 1999–2000 academic year, Massachusetts required statewide learning standards and assessment in core subjects (White, 1997). The chief education officer in Massachusetts is currently arguing for a mandatory high-school graduation examination analogous to the GED. New Jersey is trying to link funding levels for schools to statewide academic standards and their assessment (Johnson, 1997b). In place of the previously used Metropolitan Achievement Tests, Rhode Island is now using criterion-referenced tests in selected academic subjects to measure how students do when compared against a state goal for performance (Archer, 1997). In these new criterion-referenced tests, the state has defined how good is good enough instead of using national norms on specific standardized instruments. Rhode Island's approach to assessment, particularly in math achievement, is related to the activities of the New Standards Project, a collaborative effort of more than a dozen states that developed standards and related assessments for student performance as benchmarked against national and international standards of what students should know and should be able to do.

Texas has developed the Texas Assessment of Academic Skills, which uses standardized tests as gatekeepers to high-school graduation (Lawton, 1997a). North Carolina has intensified its focus on teaching a state curriculum around which assessments are designed to hold schools and school districts accountable for student achievement (Manzo, 1997). Michigan has passed a bill to revise its high-school testing program to grant "state endorsements" in math, science, social studies, and communication arts (Johnson, 1997a). These endorsements are graded by student performance level—basic, above average, or outstanding—and appear on transcripts rather than on diplomas.

Proposed new rules under IDEA (the Individuals With Disabilities Education Act) provide new requirements for the individualized education plan that each student with a disability must have and for the inclusion of these students in academic assessments (Sack, 1997). These rules require that all students with disabilities must either be included in state or district assessments or be given an alternative examination. Further, these rules mandate that states must set performance standards for students receiving special education services that are similar to those for students without disabilities.

Other state and federal policies could be identified here to illustrate the comprehensiveness with which federal and state policy now treats academic standards, accountability, and assessment as interactive. But it also needs to be noted that such interactions are not moving forward without challenge. For example, the Texas Exit Exam has been the focus of lawsuits because the passage rates of African Americans and Latinos in that state are significantly below that of White students (Lawton 1997a). The North Carolina assessments are being challenged by parents and some testing experts who argue that the North Carolina end-of-grade tests were designed to hold schools and school districts accountable but are instead being used for individual assessment to determine which students should be held back or promoted or receive remediation. The Michigan tests are the focus of parental objections because they are too time-consuming (they were taking nine hours but are being shortened to six hours) and they inappropriately claim students do poorly on tests.

The National Association of State Boards of Education has contended that, "Assessments must be in tune with rigorous state standards, address specific goals, offer some national and international comparisons, include all students, and be thoroughly evaluated. . . . In this view, an effective assessment system should also help a state identify learning groups and high achievement. . . . The state then has the obligation to follow up, providing help to the students who still need to meet academic goals or to guide the offering of more instruction to foster continued achievement in the most accomplished of students" (Lawton, 1997b, p. 7). However, the National Association of State Boards of Education also provides the following caveats: "Denying a diploma based only on test scores when the student is otherwise qualified to graduate means that students who do well in school but perform poorly on the state assessment may be unfairly penalized by a one-shot evaluation of their accumulated school work" (Lawton, 1997b, p. 7).

However you define these issues and challenges, assessment has become a high-stakes mechanism affecting the life chances of many young people and substantially defining the curriculum to which teachers will teach in order to have their students perform as well as possible in state assessments to which they are exposed. Currently, 26 or more states rely entirely or almost entirely on multiple-choice tests to measure students' knowledge and skills in all subjects.

Embedded in these trends at federal and state levels are a hotbed of issues and challenges that either revisit continuing and recurring questions or identify emerging and future challenges. These include issues that focus on the tests as emblematic of potential federal control over all subject matters, of a potential national and centralized curriculum, of the inappropriateness of the proposed tests to identify individual deficits in the academic areas being assessed, and of the use of national tests presented in English to measure reading and mathematics of children whose first language is Spanish or Chinese, as occurs in specific school districts around the country (for example, California and Texas).

Among the major issues that continue to ferment are explicit or implicit concerns about test bias and gender differences. One new perspective on these issues comes from studies by Supovitz and his colleagues which contend that using standardized tests with a multiple-choice format as the predominant form of assessing the achievement of America's children is inappropriate. Supovitz (1997) argues that a diverse society deserves a more diverse assessment system. He contends:

> Of course standardized tests are biased. But it is not just standardized tests—any simple testing method is biased because it applies just one approach to getting at student knowledge and achievement. Any single testing method has its own particular set of blinders. Since the bias in testing is intrinsic in the form of assessment used, we cannot eliminate this problem simply by changing the questions asked. Rather we must ask the questions in many different ways. . . . Today's large-scale, largely multiple-choice assessments exist in a vacuum. They stand alone, inflating their importance. Since there are no forms of assessment that, in combination with standardized tests, can provide a more robust image of a student's capabilities, we have come to rely on one particular type of assessment as *the* measure of student achievement. Standardized tests are the only game in town. . . . What we need are more experiments employing combinations of assessment approaches to arrive at an appropriate melding of test forms both economically feasible and robust enough to minimize the bias inherent in any single measure alone. . . . In the end, the larger, more intractable sources of disparities in student performance stem from broad social and educational inequities. But within the realm of assessment, the challenge for educators and policy makers is to find the appropriate balance of a variety of assessment forms, so that students of different genders, from different backgrounds, and with different affinities can demonstrate their capabilities. (pp. 34, 37)

Supovitz's perspectives about test bias lead to some related perspectives that are inherent if not explicit in the debates about specific uses of assessment by parents, minority groups, politicians and, indeed, testing experts. From a multicultural perspective, many observers have a continuing concern that testing is sexually or racially biased and, indeed, penalizes rather than facilitates the growth of specific groups of clients. Although some of Supovitz's recommendations would be helpful in ameliorating such matters, still others argue that the reasons for testing have changed during this century and that purposes and uses of testing and assessment must change accordingly. As I read of the debates in Washington and in the states about assessment, I am frequently reminded of the important insights of Gordon and Terrell (1981) two decades ago. They stated, "Critics of testing argue from a sociopolitical context,-and thus challenge the very purpose as well as the developed technology of standardized testing. Defenders of testing argue from a traditional psychometric context, with little or no concern for political and social issues. The arguments of the two parties cannot be understood and appreciated without reference to those contexts" (p. 1167).

Gordon and Terrell argued that the reasons for testing have changed since the beginning of the twentieth century and several decades afterward, and so must the nature of testing. The purpose of testing has evolved from a goal of a meritocratic selection of a few who would be allocated special opportunities has given way in response to changes in the social and political environment. The assertion of group superiority on the basis of test scores and the subsequent control of the opportunity and reward structure to maintain low-status groups in some socially assigned position into an attempt to democratize access to opportunity; thus, the use of tests also should change. As understanding grows about the pluralism in and diversity of the effects of ethnicity, sex, race, and social class upon cognitive and affective structures, learning styles, motivation, and related matters, these should be reflected in purposes for assessment. In a sort of precursor to the current perspectives of the IDEA about testing, Gordon and Terrell (1981) also stated:

> The proper course of assessment in the present age is not merely to categorize an individual in terms of current functioning, but also to describe the processes by which learning faculty and disability proceed in a given individual so that it is possible to prescribe developmental treatment if necessary. . . . The equalization of opportunity may require that intervention be responsive to the functional characteristics of the person to whom the opportunity is being made available. It must be determined where the examinee is in terms of function, how he or she got there, and how growth within the examinee's particular social and cultural environment can be enhanced. (p. 1170)

Former New York Governor Mario Cuomo, in a speech to the Council of the great City Schools' annual conference, made the same points succinctly in relation to standards and assessment, "If we're going to set the bar high, we're going to have to have all the things we need to get the children over that bar" (Reinhard, 1997). Such an issue is relevant not only to children but to adolescents and adults as well.

Underlying the perennial debates about assessment and the use of tests for accountability and other purposes is the not always well-articulated reality that any test, assessment, or other measurement procedure has many validities, not one (Messick, 1995). In fact, researchers, policymakers, teachers, or counselors must be concerned not only about the validity of the measure itself but also about the validity of the inferences drawn from the measure. Thus, no matter how scientific or empirical the development of any measurement instrument may be, its probable multiple validities and the inferences that can be made from it bring both the test and the inferences into the area of values and social contexts. Frequently, then, those who argue for or against tests are really arguing about the different validities or inferences that can be assigned to these tests. Whether or not we recognize it, many of the controversies about standardized tests and other forms of assessment can be dismantled into issues that have to do with the constructive or predictive validity of tests on the one hand, and such issues as the utility of test information, or perhaps more precisely, the social functions of standardized tests on the other.

Although I have lingered on current federal policy in education as a hotbed for questions and challenges to assessment, many of the same concerns apply to the use of testing and assessment in other contexts, such as universities and workplaces, and also raise additional questions of particular relevance to counselors. Before I turn to these issues briefly, let me raise a number of questions that evolve from the observations I have made so far:

1. How do we develop the infrastructure to create the assessment strategies necessary for a diverse society? Such a society merits the investment of resources and time to invent, test, and integrate into a program of assessment measures that go beyond standardized, multiple-choice instruments (e.g., performance measures, portfolio measures).
2. Are we fully aware that the use of assessments does not exist in a vacuum and that their purposes are shaped by changing social values as well as by their psychometric properties?
3. As a profession of persons interested in the development and use of assessments, are we fully attentive to the reality that much of the debate about tests is political, not scientific? Do we have the will and the insights to enter that debate and bridge the often disparate voices on either side of the technical-social validity debate?
4. Do we have the technical capacity, the researchers, and the test developers to meet the challenges created by the growing expectations that assessments of different kinds will be increasingly central to matters of school reform and other sociopolitical purposes?
5. Should we be advocating national standards and constructing specific assessments to evaluate whether they are achieved? If so, how do we integrate the scoring and interpretation of these

assessments with advanced technologies? What trade-offs between centralized and decentralized approaches to academic standards, to forms of assessment, to norms, and to multicultural issues are we as assessment professionals prepared to accept?

6. Are we training teachers, counselors, and administrators to view assessments in comprehensive terms? Are we helping them understand the political climate for assessment and the validities inherent in using assessment devices?
7. Are we prepared to respond to questions from the media and from policymakers with regard to whether national testing and national standards are good ideas?
8. Within basic education settings, are school counselors acting as resource persons or implementers of assessments for accountability, for exit exams, or for other purposes? Are they being trained to play these roles?

Although many of the questions that arise from current national and state debates about assessment in education are the content of media headlines, many less publicized but similar questions arise as assessments are being applied to various groups of adolescents and adults, such as those moving from school to work or from welfare to work, dislocated homemakers or women attempting to reenter paid employment, and military applicants in an increasingly technological environment and one that is changing rapidly in the proportion of males and females. Assessments, including literacy audits, are being applied to current members of the American workforce to measure their competencies; basic academic skills; and ability to learn new industrial, manufacturing, or business processes. We also need to ensure that people with disabilities are able to use their talents and skills in educational and work settings without discrimination and bias. Finally, immigrant populations and cross-national populations need to be assessed when they are assimilated into or recruited for American jobs for which there are skill shortages.

Embedded in such adult employment initiatives are both implicit and explicit expectations that various types of assessment will be important, and there are many questions to be answered. For example, "What workplace skills are to be assessed?" There is currently great interest in assessing "soft" skills such as interpersonal skills and work habits, as well as traditional cognitive skills. Other assessment questions raised at the national level relate to the use of ethnographic approaches to study the use of literacy in the workplace, the importance of informal knowledge for gaining vital on-the-job skills, the use of assessment-center procedures developed by AT&T 40 years ago for measuring job skill attributes required for the twenty-first century, and the use of video, computers, or the World Wide Web to overcome the performance barriers many minority persons experience in taking written tests. Such proposed directions for assessment relevant to the recruitment, induction, and retraining of individuals in the workplace keep pushing the envelope of available research on such processes and exposing the need for new mindsets and initiatives in assessment.

Although space does not permit an extended analysis of the issues and challenges that relate to the assessment of these adult populations, suffice it to say that among the issues are the use of standardized, multiple-choice, knowledge-based tests versus performance-based assessment; new forms of functional analysis for persons with disabilities; the gender and racial biases of norms; a lack of knowledge about men and women with low socioeconomic status in terms of their characteristics and lived experiences, their learning styles, their inexperience with assessment processes, and how these factors affect their scores. In addition, as a more pluralistic and culturally diverse population certainly translates into a more culturally diverse workforce, how do we create a diverse assessment system for adults that accommodates language differences and differences in educational backgrounds in the countries of origin from which immigrants are coming? How do we incorporate responses to these issues into tools for employment and career counselors, rehabilitation counselors, military classification experts, and counselors in educational settings?

Given the limitations of space, then, let me turn in the remainder of my article to the historic interaction between counseling and assessment, testing and assessment as interventions, and how important assessment has been and is in bridging the gap between theory and practice. Then I will pose some questions about such matters as well.

The alliances between counseling and assessment have ebbed and flowed depending upon what counseling and personality theories were in vogue at particular times in our history, the types of training provided counselors at different points in the past century, and the degree to which assessment has been seen as a legitimate and useful adjunctive input or complement to counseling. The use of testing and assessment by counselors has been affected by a number of issues of gender or racial bias of tests, as well as other matters that have already been discussed. In any case, there are a number of issues affecting assessment in counseling in schools and in other settings that are hotbed issues and challenges. These issues have to do with a range of process concerns: Are assessment processes being used effectively in counseling? Are assessments really interventions in their own right? Do assessments effectively bridge the gap between theory and practice? There are also a number of professional issues: Who should test? Are counselors and therapists being effectively trained to test, and how do we know that this is true?

Although it is tempting to go back to the beginning of the twentieth century and trace the important interaction of counseling and assessment as both have grown in maturity during the century, I will resist that urge. Instead let me suggest briefly that to a large degree changes in counseling and in assessment have frequently coincided with emerging theories of life span psychological development; client-centered or cognitive-behavioral counseling; and, particularly, the expanding models of career development of persons such as John Holland (1992), Donald Super (1990), John Krumboltz (1994), and many others, as well as the attempts of these theorists to make their theories accessible to counselors through the use of assessment instruments. Indeed, given these circumstances, I am frequently puzzled at the continuing criticisms by counselors and by some counselor educators that formal personality, counseling, or career theory is not relevant to what counselors do, or that theory and practice are separated because theorists do not tell counselors how to use their theories. I respectfully suggest that in large part that criticism is a myth rather than a reality. Let me take career-development theory as an example. In my view, assessment has been the bridge in operationalizing theoretical constructs by reflecting them in interventions and, in particular, in tests and measurements.

Certainly, this has been true of Holland's theoretical constructs, which are reflected in the Vocational Preference Inventory, My Vocational Situation, and the Self-Directed Search; in the use of his theoretical framework (RIASEC) as the organizing and interpretive structure for the most recent interactions of the Strong Interest Inventory and for some of the informational and self-assessment components of the DISCOVER computer-mediated career-guidance system; and in the use of Holland's three-letter coding system of major personality types as a way of organizing U.S. government educational and occupational information through such sources as the *Dictionary of Holland Codes*.

Similarly Super, from the beginning of his conceptual work, has used assessment instruments to operationalize and to evaluate his theoretical constructs. Like Holland, he has made his theoretical constructs accessible to practitioners by using assessment to bridge theory and practice. Relevant examples include the Career Development Inventory, the Adult Career Concerns Inventory, the Work Values Inventory, and more recently, the Values Inventory, the Salience Inventory, and the Career Rainbow. Each of these instruments attempts to describe or to measure individual career behavior in ways that are useful in defining goals for counseling and in explicating one's maturity or one's levels of career planning, knowledge and attitudes about career choice, intrinsic and extrinsic life-career values, and the relative importance of major life roles beyond those of occupation or career.

Super's theoretical work has spawned (a) assessments by others (for example, Crites' Career Maturity Inventory); (b) theoretical extensions like that of Gottfredson's (1981) processes of *circumscription* and *compromise* as ways of incorporating the effects of gender issues, sex bias, and sex roles as factors shaping the roles of women; and (c) indeed, models of career counseling, like the recent C-DAC in which assessment and counseling are intimately interactive as interventions.

Although there are many other examples, in and out of career theory, of instances in which assessment has been used to bridge theory and practice and, indeed, has been conceived of as an intervention in its own right, I will conclude by briefly acknowledging the importance of John Krumboltz's theoretical concepts through the years, his development of innovative assessment devices during his earlier emphases on behaviorism and as he has articulated his social learning theory and, more recently, his cognitive-behavioral theory related to such issues as faulty self-observation generalizations or inaccurate interpretations of environmental conditions, and his recent development of the Career Beliefs Inventory as a counseling tool by which to identify presuppositions and irrational beliefs that may block people from achieving their goals. I must confess that I sometimes wonder whether counselors are being trained to understand the intimacy of theory and the assessment instruments which have been derived from theory and which serve to stimulate client self-appraisal as an important intervention alone and as a stimulus to creating the content which counseling explores, clarifies, and incorporates into individual plans of action.

Now let me review quickly some of the trends of the late 1990s and some of the challenges for assessment that are spawned by the evolution of counseling in the United States and elsewhere. They include:

- Growing acceptance of counseling programs as central to the mission of schools, higher education institutions, and increasingly to workplaces, rather than as frills or ancillary services. In these contexts, assessment and evaluation issues are increasingly seen as major tools pertinent to the integration of institutional missions and the deployment of counseling resources and purposes.
- The systematic development, planning, implementation, and evaluation of counseling programs in schools, colleges and universities, and workplaces. Such programs are increasingly seen as having their own psychosocial content (e.g., career planning, purposefulness, productivity, stress management,

anger reduction) and their own responsibility for facilitating certain types of student or client knowledge, attitudes, and behaviors. Rather than being a random collection of services or functions, or by-products of other activities, counseling programs are increasingly expected to identify the results for which they will be accountable and to provide evidence of their effectiveness for accountability purposes. Assessment will play an important role both as an intervention within such programs and as an evaluation tool.

- Issues of cost/benefit ratios relative to counseling programs increasingly will be an issue in the next century. In the United States we have rarely raised that issue, but our colleagues in Europe have and we can expect that it will be an emerging concern in schools, in workplaces, and in higher education. Assessment strategies will be critical as they relate to producing relevant measures or assessments of productivity by counselors in different environments. Other critical uses of assessment will be as measures of counselors' effectiveness in terms of the outcomes they obtain with individuals or groups of students, clients, or consumers, as well as the costs of producing the units of productivity measured in relation to, for example, the use of goal-directed, time-limited interventions versus psychoeducational models or the use of technology.
- Another issue is the use of needs assessments to identify the topics or problems counselors in different settings should be addressing and the differential treatment-by-client interactions that should be planned for in designing counseling programs. Again, needs-assessment strategies and the measures useful in comparing the effects of differential treatments for specific common outcomes will need to be refined and enhanced.
- Another pervasive theme in the next century will be attention to crisis intervention and to addressing the needs of persons at risk (e.g., those who experience chemical dependence, are violence prone, are likely to be an academic or work failure, are likely to have a teenage pregnancy, or are socially or emotionally dysfunctional). Related will be new approaches to (a) early identification, prevention, and treatment; (b) more participation of counselors in student-assistance or employee-assistance programs or other group or shared approaches to intervention for different populations and purposes; and (c) more inclusion of counseling in a total program of interventions aimed at the multiple problems experienced by most people at risk. Assessment needs will pervade such trends. They will include increased emphases on diagnosis; on identifying the

competencies that counselors need to work with these populations; and in combination with other mental health professionals, on assessment of differential individual and group treatments, their effectiveness and their cost/benefit ratios.

- We are seeing more uses of technology (e.g., computer-assisted career-guidance programs, testing and test interpretation conducted by computer, career counseling on the Internet, self-directed planning and decision making, distance learning, electronic information processes). These need to be evaluated in terms of both outcomes and differential treatment.
- In the next century, we are likely to make greater use of differentiated staffing among counselors in particular settings and of new configurations of counselors, support systems, and technology to deal with different demographic profiles and institutional emphases. Again, there will be increased needs for needs assessment, comparative analyses, cost/benefit research, and other types of assessment.
- Finally, there will be an increased emphasis on training counselors in competency-based formats, with different mixes of didactic and hands-on supervision, for work in different settings and with different populations. In these contexts, there will be increased concern about assessing counseling learning styles and preferred modes of training, about the use of virtual reality in lieu of or support of the supervised practice of counseling skills, and about the types of context-dependent assessment competencies, as they are applied manually and through technology, that counselors need.

Obviously, this litany of potential trends is in no way exhaustive, but it suggests what would appear to be a growing need for clarity about how counseling and assessment should be interactive. These trends acknowledge that external political, legal, economic, and social forces will likely modify or add to emerging trends relating to the importance and character of counseling programs.

One of the growing political and economic challenges for counselors, either directly or indirectly, is the current national rhetoric about certifying competencies. As the policymakers continue to engage in school reform, redefinition of workplace education and development, school-to-work transitions, and workplace reorganization, they will place an increasing priority on the certification of competencies possessed by school and university students and by workers. Employers are no longer satisfied to accept program completion as evidence of employability or occupational skills. Instead they expect competency certification at various levels and in different paradigms, and assessment measures will be sought to provide such certification. Given the changes in the nature of the workplace and in the skills required to work with new industrial and business processes, in technologically intensive environments, and in collaborative work groups, one can expect that certification of students' or workers' competencies will go beyond measures of competitiveness, problem-solving ability, and resemblance or similarity to work groups. Instead certification will direct greater attention to competencies underlying complementarity—the ability to facilitate the work of others and engage in group problem-solving—career motivation, career resilience, career identity, career insight, personal flexibility, and teachability.

The fundamental point here is that the application of assessments to questions of individual competence and program accountability is going to be a major issue far into the twenty-first century. Counselors are not likely to be exempt from such assessment concerns; the notion of certifying competencies will extend to them as well. Obviously, as a profession we are well along that road because of the pioneering leadership of the Association for Assessment in Counseling (AAC), the National Board of Certified Counselors (NBCC), the National Career Development Association (NCDA), the Council for Accreditation of Counseling and Related Educational Programs (CACREP), and other American Counseling Association (ACA) units. But even given the excellence of these efforts, most of the certification approaches to date have been knowledge-based not performance-based, at least as they relate directly to the impact of the counselor on clients. These issues are likely to become more delicate in the future as various mental-health professional organizations such as the APA try to define the scope of practice of their constituents—psychologists—to encompass that which counselors have historically been trained to do and have done. As you may know, counselors in California, Georgia, Indiana, Louisiana, and other states have faced challenges to their use of tests in counseling. Psychologists in these states have mounted efforts to restrict the use of tests to doctoral-level professionals. The latter is often a code word in specific states for persons who have been trained in APA-accredited counseling psychology programs, not in doctoral programs in counselor education. The NBCC board of directors has addressed this challenge by citing a number of points that it feels are important to the assessment practice of National Certified Counselors (Clawson, 1996). They include:

- The practice of counseling requires a right to administer and interpret standardized psychometric assessment instruments (tests) to plan treatment or to assist with life planning.
- The right to administer tests should be based upon adequate training, not on academic degree or discipline.

- The current behavior of the American Psychological Association (APA), state psychology licensure boards, and state psychology associations regarding discipline "ownership" of psychological testing is improper.
- No counselor should administer any assessment instrument without proper training.
- Training institutions should instruct counselors in the proper use and awareness of testing procedures.
- Members of the counseling profession, including all counseling professors, should promote proper use of tests and advocate for counselors' right to use tests.

We must recognize that both APA's actions and NBCC's responses are political not scientific responses. Therefore, if this issue continues to ferment because of credentialing competition, its future resolution will lie in answers to more precise assessment questions: Who has the competencies that can be demonstrated to be accurate, relevant, and effective relative to clients' needs? What is adequate or proper training and how can it be assessed? What are the specific competencies that counselors-in-training achieve through test and measurement courses as part of their core preparation? How do these compare with those possessed by psychologists (at master's and doctoral levels)? How should counselors' competencies differ from those of psychologists in relation to the types of tests being used in counseling practice (e.g., the assessment and diagnosis of emotional disorders versus the assessment and diagnosis of aptitudes, interests, career maturity, etc.)? How do we insure that we recognize the key role of the test user in testing? That is, we must effectively respond to the observation of Anastasi (1992, p. 610) that "most popular criticisms of tests are clearly identifiable as criticisms of test use (or misuse), rather than the tests themselves. Tests are essentially tools. Whether any tool is an instrument of good or bad depends on how the tool is used." These questions and useful answers to them are embedded in ACA's statements of ethics and in the packets of information sent to the Attorneys General of Georgia and Indiana in response to the challenge to counselors' right to do psychological testing in these states. These questions and others that need to be more fully incorporated in counselor competency assessment are also included in such documents as *Test User Qualifications: A Data-Based Approach to Promoting Good Test Use* (Eyde, Moreland, Robertson, Primoff, & Most, 1988, p. 14), which includes a factor analysis of good testing practices that yielded seven tentative factors and recommendations for fundamental operating principles:

1. Comprehensive assessment	Following up to get facts from psychosocial history to integrate with test scores, as part of interpretation
2. Proper test use	Accepting responsibility for competent use of the test
3. Psychometric knowledge	Considering standard error of measurement
4. Maintaining integrity of test	Making clear that cutoff scores imposed for results placement in special programs for the gifted are questionable because they disregard measurement error
5. Accuracy of scoring	Using checks on scoring accuracy
6. Appropriate use of norms	Not assuming that a norm for a given job (or group) applies to a different job (or group)
7. Interpretive feedback	Willingness and ability to give interpretation and guidance to test takers

Fundamental operating principles that guided this data-gathering effort were:

1. A model test-user qualification system should be based on scientific methods and should serve as a tool for identifying the competencies of test users.
2. Access to psychometric instruments should be based on knowledge and behavior of test users rather than solely on job titles or credentials.
3. The key to the model system is self-regulation.
4. The model applies to a broad range of test users who belong to many different professional associations that engage in professional self-regulation, using ethical principles relating to competence.
5. Legislation restricting test use to psychologists or psychologists supervised by psychologists is unrealistic and unnecessarily restrictive, and applies primarily to tests used by psychologists, thus ignoring other practitioners.
6. Test misuse is more likely to be a function of lack of information or misinformation than of malfeasance on the part of the test user.
7. Educational efforts are likely to be more effective than

restriction of access in promoting good testing practices.

8. The proposed competency-based user-qualifications system, which is designed to reduce test misuse, is likely to increase the use of tests as an important element in decision making.
9. By identifying possible test misuse, the system will alert test users to poor testing practices and reduce the likelihood that tests will be banned through legislative action.

This document was derived from an interdisciplinary model that represents an alliance of organizations whose practitioners engage in assessment; it is an important and enlightened approach to the credentialing mania and turfdom that is again arising. Interestingly, the operating principles embodied in it are essentially the same as those included in the American Counseling Association ethical standards (as revised in April 1995) and they are consistent with the counselor preparation standards promulgated by CACREP.

Unfortunately, as I have said earlier, each of us must recognize that this growing challenge to counselors' use of assessment by psychologists is not a matter of science or of aggregated research findings but instead of power, protection of the independent marketplace, and politics. As issues of power and politics continue to arise about who should be able to test and within what scope of work, one of the facts that helps to explain the rising tensions among professional groups is that tests, for reasons I have already addressed, have become terribly important elements in contemporary society. They are important in the accuracy or inaccuracy of their content in relation to their purposes; they are important in their application and interpretation; they are important in their uses for classification, inclusion, or exclusion; and they are important in the populations for which they are relevant. Their development and use are also worldwide. Indeed, a substantial number of recent conferences held in Greece, Spain, Germany, and Belgium have focused on the use of psychological assessment. Recently, Division 2 of the International Association of Applied Psychology has begun the process of developing Guidelines for Adopting Psychological Tests for Use in Multiple Languages and Cultures. Underlying the development of these standards are the increasing use of tests cross-culturally and questions of the validity of such uses. *Multicultural Assessment Standards: A Compilation for Counselors*, edited by Dale Prediger (1993), is an excellent reference that puts many of these issues into context.

Much more could be said about the challenges just cited, but I will move on to a brief mention of some final assessment issues. The first is teaching the test. Certainly the use of tests as diagnostic instruments to identify developmental deficits or psychological traits or states has a long and important history. However, we often consider all of the scores from these assessments as fixed effects rather than recognizing that some cognitive or behavioral areas are fixed effects but other areas are more malleable individual characteristics that are susceptible to learning on the part of the individual (in particular by teaching persons why their answers on a certain test were wrong and what is implied for them in learning or relearning certain types of behavior or knowledge). In such cases, depending upon their uses, tests can be interventions. This is a different mentality about testing than the mindset that the scores attained are absolute and not susceptible to modification. One can argue that coaching people to do better on the SAT, the MCAT, or various occupational entrance examinations is teaching the test, and that the data are mixed about whether such coaching does any good. True enough! But the response depends in part upon what tests one is talking about (e.g., career development process instruments, etc.) and how one views related observations by Charles Healy and others that focus on helping clients to develop self-assessment skills and to become true collaborators in the appraisal process. In particular, Healy (1990) has talked about *four obstacles* to such a shift in thinking about new counselor-client collaborative models of using appraisal data in career counseling. We have all heard them many times: "(a) casting clients as subordinates rather than as collaborators; (b) discounting self-assessment by favoring counselor assessments; (c) de-emphasizing the influence of contexts in client development; and (d) focusing on a single choice rather than on strengthening client decision making and knowledge for follow-through" (p. 214). Views about teaching the test are obviously impediments to the growing need to empower clients by strengthening their assessment skills through teaching the test and using the content to encourage client self-evaluation and decision making. Prediger (1993), Zytowski (1994), and Kapes, Mastie, and Whitfield (1994) have also discussed ways in which the relationship between testing and counseling can be enhanced, not fragmented.

A further challenge, although not necessarily a new one, has to do with computer applications in testing. In one sense such applications have become commonplace, but they are also uneven in their use across settings, populations, geographical regions, etc. Computer applications in testing include the self-assessments embedded in computer-assisted career-guidance programs, but they extend also to the administration, scoring, and interpretation of tests. We are finding increased use of computers for self-help programs of all kinds, including those purporting to provide personal counseling, in the form of expert systems modeling counselor behavior in responding to a client's descriptions to the computer of his or her psychological dilemmas. The computer is also being used by some health personnel for consultation in emotional crises where psychiatrists, psychologists, or counselors are not immediately available.

The fundamental point is that the use of computers in testing, in statistical analysis, and in all sorts of other related ways sometimes occurs in immediate conjunction with the process of counseling or psychotherapy, not simply as administrative procedures unconnected to counseling. As a result, there are continuing and in some ways increasingly complex ethical questions involved in the application of computers to testing, to self-appraisal, to personal counseling and to the variations on these themes.

Computer-based test administration and interpretation, like every other technique available to a counselor, can be both a boon and a bane. On the positive side, it can be cost-effective and, in the case of microcomputers, can provide test information virtually instantaneously. In general, clients seem to enjoy the experience and to achieve as much self-knowledge as when paper-and-pencil tests are used. Further, no violence seems to be visited on the psychometric properties of accepted testing instruments that are computerized (i.e., validity, reliability, etc.). The negatives of computer-based testing are involved more with the idiosyncratic aspects of a particular instrument, interpretive program, or hardware configuration than with the concept itself. Group administration is obviously difficult, if not impossible, because of the prohibitive cost of multiple stations; some programs are not user-friendly; some instruments are so new and are rushed to market so quickly that they provide inadequate validity and normative data; and erroneous or overly generalized interpretations are possible. Further, in reality there is as yet little research to determine individual differences in person-machine interactions. A final concern is perhaps the most ominous—counselors may believe that because the machine is producing an impressive-looking report, they need not have an in-depth knowledge of the test, its underlying constructs, its psychometric strengths and weaknesses, appropriate interpretations, and the need to integrate the results with all other information relevant and important in the career development of the client.

I would be remiss here if I did not mention the challenges to assessment and to the ethics of assessment that are now inherent in the Internet. I do not have to remind you that as a nation we have embraced the Internet with a passion that belies the reality that there has been virtually no research done about the effects of the Internet on learning, mental health, career decision-making, and so on. Some 50,000 to 60,000 pages are being added to the Internet each day, and much of this content purports to be relevant to what counselors do.

Inherent in the Internet is concern about ethical research in the information age. The implication is that researchers who study electronic communities or on-line communities will likely find themselves increasingly using qualitative methods, changing their commonly used research tools, and adapting their assessment strategies to these new electronic environments. In essence, each of the current capabilities of the Internet, from e-mail to chat rooms, will pose its own research and assessment dilemmas related to how to obtain informed consent; how to conceive of respondents as owners of the materials they create; how to protect copywritten material on the Internet; how to create a climate of trust, collaboration, and equality with electronic community members; how to negotiate researcher entrée into an electronic community; how to treat electronic mail as private correspondence, not to be treated as research data unless express permission is given; how to respect the identity of the research respondents in an electronic community, to protect or mask the origins of the communications, and to communicate the results of their research to participants in the research (Schrum, 1995).

In conclusion, then, let me say what you already know—there is much more to say and challenges and issues that are yet to be identified relating to assessment in the twenty-first century. What is apparent in these deliberations is the reality that during the twentieth century, both assessment and counseling sank their roots deep into the American social fabric, and both have matured in their conceptual and methodological processes. Both will be extremely important in the twenty-first century as they contribute to national goals of mental health, career development, productivity, and individual purpose. But meeting these goals continues to raise questions that must be addressed systematically and scientifically. They include:

1. Are we training counselors in the most effective ways to use assessments, and to understand the roles of assessment as interventions and as integral to counseling processes?
2. Have we identified effective training models in counselor use of assessments and the competencies necessary to use different types of tests in counseling practice? Are we providing sufficient opportunities for retraining of counselors whose skills and understanding of assessment may be outdated?
3. Are we training counselors to use assessments in new and emerging contexts: to teach the test, to use computer-assisted test interpretations, and to use the World Wide Web to do assessment?
4. Have we considered how different groups of helping professionals differ in their ability to use assessments and how they might complement each other in school, community, or workplace contexts?
5. As specialists or users of assessments, are we prepared to

understand the political as well as the scientific and technical issues related to the uses of assessment in counseling?

6. Are we prepared to talk about the cost-benefit ratios of assessment used in different forms and models and in relation to different models of counseling practice? Are we prepared to talk about the assessment of counseling in terms of both productivity and effectiveness?
7. Are we adequately preparing counselors to think and act in multicultural terms as they address assessment issues? Do counselors understand and act in accord with existing research showing that persons from different national and cultural traditions—even those who are residents of a pluralistic nation such as the United States—may have different values, beliefs, communication styles, methods of solving problems, perceptions about problems, and ways of coping with them (Wilgosh & Gibson, 1994)?

Although not exhaustive of the questions before us, this list, like the issues inventoried earlier in this article, are representative of the issues and challenges that await us in the new millennium.

References

American Counseling Association. (1995). *Code of ethics and standards of practice*. Alexandria, VA: Author.

Anastasi, A. (1992). What counselors should know about the use and interpretation of psychological tests. *Journal of Counseling and Development, 70*(5), 610–615.

Archer, J. (1997, October 29). New test in r.l. sends shock waves through state. *Education Week*, 15.

Clawson, T. (1996, Winter). Fair access in testing. *NBCC News Notes, 12*(3), 1, 3.

Eyde, L. D., Moreland, K. L., Robertson, G. J., Primoff, E. S., & Most, R. B. (1988). Test user qualifications: A data-based approach to promoting good test use. *Issues in scientific psychology*. Washington, D.C.: American Psychological Association, Science Directorate.

Gordon, E. W., & Terrell, M. D. (1981). The changed social context of testing. *American Psychologist, 36* (10), 1167–1171.

Gottfredson, L. S., (1981). Circumscription and compromise: A developmental theory of occupational aspirations. *Journal of Counseling Psychology, 28*(6), 545–579.

Healy, C. C. (1990). Reforming career appraisals to meet the needs of clients in the 1990s. *The Counseling Psychologist, 18*, 214–226.

Hoff, J. (1997, November 5). Latest testing compromise appears doomed. *Education Week*, 22, 26.

Holland, J. L. (1992). *Making vocational choices* (2nd ed.). Odessa, FL: Psychological Assessment Resources.

Johnson, R. C. (1997a, October 22). Michigan House passes bill be revise high school testing program. *Education Week*, 11.

Johnson, R. C. (1997b, December 3). Education prominent in N.J. governor's race. *Education Week*, 15, 17.

Kapes, J. T., Mastie, M. M., & Whitfield, E. A. (1994). *A counselor's guide to career assessment instruments* (3rd ed.). Alexandria, VA: National Career Development Association.

Krumboltz, J. D. (1994). Improving career development theory from a social learning perspective. In M. L. Savickas & R. W. Lent (Eds.), *Convergence in career development theories. Implications for science and practice* (pp. 9–31). Palo Alto, CA: CPP Books.

Lawton, M. (1997a, October 22). Discrimination claimed in Texas exit-exam lawsuit. *Education Week*, 3.

Lawton, M. (1997b, December 3). State board leaders call for assessments bearing consequences. *Education Week*, 7.

Manzo, K. K. (1997, October 22). High stakes: Test truths or consequences. *Education Week*, 1, 9.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist, 50*, 741–749.

Prediger, D. J. (Ed.). (1993). *Multicultural assessment standards: A compilation for counselors*. Alexandria, VA: Association for Assessment in Counseling.

Reinhard, B. (1997, October 22). Reporter's notebook. *Education Week*, 5.

Sack, J. (1997, October 29). Proposed IDEA rules target testing, accountability. *Education Week*, 19, 24.

Schrum, L. (1995). Framing the debate: Ethical research in the information age. *Qualitative Inquiry, 1*(3), 311–326.

Super, D. E. (1990). A life-span, life-space approach to career development. In D. Brown & L. Brook (Eds.), *Career choice and development. Applying contemporary theories to practice* (pp. 197–261). San Francisco: Jossey-Bass.

Supovitz, J. A. (1997, November 5). From multiple choice to multiple choices: A diverse society deserves a more diverse assessment system. *Education Week,* 34, 37.

Terwilliger, J. (1997). Semantics, psychometrics, and assessment reform: A close look at 'authentic' assessments. *Educational Researcher, 26*(8), 24–27.

White, K. A. (1997, November 5). Duty-bound. *Education Week,* 28–32.

Wilgosh, L., & Gibson, J. T. (1994). Cross-national perspectives on the role of assessment in counselling: A preliminary report. *International Journal for the Advancement of Counselling, 17,* 59–70.

Zytowski, D. G. (1994). Tests and counseling: We are still married and living in discriminant analysis. *Measurement and Evaluation of Counseling and Development, 26,* 219–223.

About the Author

Edwin L. Herr is distinguished professor of education (counselor education and counseling psychology), associate dean for graduate studies, research, and faculty development, and interim department head for adult education, instructional systems, and workforce education and development in the College of Education at The Pennsylvania State University. A national certified counselor, national certified career counselor, and licensed psychologist in Pennsylvania, Herr has a part-time practice in career counseling and has consulted widely in the United States and abroad with government agencies, schools, universities, and corporations. Herr is past president of the American Association for Counseling and Development, past president of the National Vocational Guidance Association, and past president of the Association for Counselor Education and Supervision. He also served as President of Chi Sigma Iota in 2000-2001. He is the author or co-author of more than 275 articles and book chapters and 31 books and monographs.

1. This article was originally presented at the Assessment '98: Assessment for Change—Changes in Assessment conference, St. Petersburg, Fla., January 16–18, 1998.

Chapter Two

An Emerging Paradigm of Testing

Lorin Letendre[1]

"You must be the change you wish to see in the world."
—Mahatma Gandhi

Abstract

This article presents an emerging approach to testing that has the potential to respond to the concerns of critics of standardized testing. This emerging approach differs from the traditional, or establishment, approach in how it views human nature and the test taker, the purpose of testing, the testing process, outcomes of testing, and who makes testing decisions. The Myers-Briggs Type Indicator and the Herrmann Brain Dominance Instrument are used to exemplify how the emerging and establishment approaches differ.

Let me move directly to the central thesis of this article, which is that the limits to the growth of testing and assessment, and to our overall testing market, stem more from our own attitudes in the test-development business than from the efforts of our critics or anti-testing opponents. If we are willing to change some of those attitudes and their attendant practices, we can expand the potential for testing and assessment, and thus our potential market, far beyond its present size. The attitudes that limit our growth potential are primarily elitist, despite being clothed in scientific respectability. Those attitudes both delimit and threaten our future, and only a paradigm shift is likely to change them.

I'd like to begin developing this thesis by citing a couple of surveys of attitudes toward testing among Americans. The first is an innovative Gallup survey conducted back in 1979, in which the "Gallup pollsters asked a national sample of American parents what they thought of standardized tests" (Nevo & Jager, 1993, p. 7). Before I give you the results, I want you to guess what percentage of the Americans surveyed believed that standardized tests were "very useful" or "somewhat useful," versus what percentage believed that tests were "not very useful." Here are your choices:

	Very Useful or Somewhat Useful	Not Very Useful
A.	33%	67%
B.	67%	33%
C.	83%	17%

You might be surprised to learn that the correct answer is *C*.

Now consider the second survey, which is a more recent survey of even more laypeople. Conducted between 1980 and 1986 by the King County Civil Service Commission, it asked 2,500 job applicants who were applying for both entry-level jobs and promotions to higher jobs whether or not they viewed the standardized employment test as "fair." Now that you know the results of the Gallup poll, no doubt you will guess more accurately the result of this much more extensive survey. Here are your choices:

	Employment Test Is Fair	Employment Test Is Unfair
A.	5%	95%
B.	50%	50%
C.	95%	5%

The correct answer is *C*, which flies in the face of what many of us in the testing industry perceive as decidedly negative public attitudes toward testing and certainly toward employment testing, which potentially has a huge impact on laypeople in terms of both money and status. Although they may or may not be representative of Americans' opinions toward testing in general, these two surveys suggest that public perceptions of testing are far more positive than we inside the testing industry perceive, as judged by all our efforts to address public criticisms of testing and assessment.

I'd like to define the term *paradigm,* which I use in my title. I am suggesting that the testing industry is undergoing a gradual paradigm shift from one set of attitudes toward a new set that could resolve many of the criticisms we have faced as an industry. Thomas Kuhn, who was one of the pioneers in writing about paradigm shifts, argues persuasively that the word *paradigm* is less clear and useful than the term *disciplinary matrix,* so I am opting to use his term. Kuhn defined *disciplinary matrix* as "a complex of generalizations, beliefs, values, and exemplars that direct the normal day-to-day activities of a given scientific group." Tim Rogers applied this term to test developers and asserts that "the ideas of validity, reliability, and utility; the technologies of test construction; the prevailing ethical standards—all are part of this matrix for the testing community" (Rogers, 1995, p. 768).

You may have recognized that there is a paradox inherent in the criticisms of standardized psychological tests. Tests—particularly tests developed in ways that minimize bias—continue to be the most objective means of making high-stakes decisions, yet their visibility in such decisions and the fact that some test takers must lose in the decision-making process make tests vulnerable to criticism as the "bearers of bad news." Franklin Jenifer stated this case quite cogently: "I believe standardized achievement tests are not the problem; they serve only as the messenger bearing ill tidings. Instead of attacking and even attempting to 'kill' the messenger for bringing us news we do not want to hear or accept, it would be better to devote our attention to the message and, even more important, to the reasons why it is so bad" (Daves, 1984, p. 97).

This argument goes to the heart of our democratic belief system in America. One of the forces that underlies the criticism of standardized tests is egalitarianism, for the egalitarian complaint is that the tests discriminate among test takers and favor those with the best education and the most ability. But the force that makes standardized testing an omnipresent feature of our society is also egalitarianism, because "testing continues to be the most objective mechanism available to allocate benefits" (Daves, 1984, p. 59).

Tests are one of the few decision-making tools that has the potential to be blind to gender, ethnicity, sexual preference, age, religion, status, and other divisions in our society.

Diane Ravitch eloquently summarized another crucial point: "My own view is that the tests have become increasingly controversial because they have become increasingly indispensable" (Daves, 1984). Revulsion against standardized testing has come at a time when tests have become a fixture not only in educational decision making but in entry into the labor market. In researching this article, I considered both the critics' arguments and the rebuttals from testing experts and other professionals in the testing industry. My own conclusion is that we can have it both ways—we can continue to provide the most objective possible decision-making tools for high-stakes decisions in our economy and society, *and* we can nullify most of our critics' attacks. To do so, we must shift some of our resources to an emerging form of testing and assessment, which following Thomas Kuhn's terminology, I will name the *emergent matrix* or *emergent approach* (Rogers, 1995). I will contrast this new approach with what I will term the *established matrix* to testing, which has dominated our testing practices and unduly influenced our attitudes so as to bring down on our own heads many deserved as well as undeserved criticisms.

Our critics have argued that many mental and other ability tests have "labeled" or "diagnosed" the deficiencies or deficits of test takers, in order to justify their exclusion from educational or employment avenues down which other Americans have progressed. Stephen Jay

Gould (1996, p. 50) poignantly stated this case: "We pass through this world but once. Few tragedies can be more extensive than the stunting of life, few injustices deeper than the denial of an opportunity to strive or even to hope, by a limit imposed from without, but falsely identified as lying within."

My proposed new emergent model of testing and assessment can make a central contribution in blunting such damning criticisms. Let me outline the key features of this model, which I will do by contrasting its features with those of the established matrix. I use the term *established* for two reasons: first because over the past 90 years or so it has become a well-entrenched approach to testing, and second because it connotes the "Establishment," or the socioeconomic-political elite who occupy the top rung on the prestige ladder in the United States today. Table 2.1 compares the features of the two models.

Table 2.1. Comparison of Emergent and Established Matrices

Emergent Matrix	Established Matrix
View of Human Nature	
Emphasis on people's positive characteristics	Emphasis on people's deficits and potential deficiencies and limitations
View of the Person	
Holistic, integrative, people viewed as dynamic "wholes" or "systems"	Additive, people viewed as static mix of traits and abilities
View of the Purpose of Testing	
Serves the individual	Serves institutions
View of the Testing Process	
Open, communicative process	Closed, secretive process
View of Outcomes	
Test scores show "either-or" preferences	Test scores are right or wrong, good or bad
Locus of Control	
Test takers are decision makers	Testing professionals are decision makers

In order to illustrate the specifics of each of these models or matrices, I plan to focus on three of the features listed in Table 2.1: view of human nature, view of outcomes, and locus of control. In my examples I will use instruments that are promising examples of the emergent matrix and which are featured in a recent *Harvard Business Review* article: the Herrmann Brain Dominance Instrument (HBDI; Kramer & Conoly, 1992) and the Myers-Briggs Type Indicator® (MBTI®; Briggs Myers 1977/1999).[2]

View of Human Nature

The established matrix focuses its measurement power on the deficits, defects, and deficiencies of people, whereas the emergent matrix helps people identify their positive characteristics and build on these strengths. The established matrix is thus much more cynical and pessimistic about human nature, while the emergent matrix is optimistic about people and their ability to change and progress.

I now turn to some real-life examples, starting with the MBTI, a widely used measure based on Jung's work on psychological types. In their book on psychological types, Roger Pearman and Sarah Albritton cite its relevance to human nature: "At a minimum, psychological type provides models for two very important insights into human nature. First is a model for understanding human differences that provides hypotheses about people different from ourselves but that doesn't value one type over another. Second is a model that provides basic questions to help us solve problems in any situation or interaction" (Pearman & Albritton, 1997, p. 164). They emphasize the centrality of healthy human development: "Psychological type and type tools like the MBTI provide a very positive and constructive model to understand differences in the way individuals process and express information. . . . While it is valuable to know one's type and what some typical reactions and blind spots may be in a particular situation, it is for developmental rather than diagnostic or managerial use" (Pearman & Albritton, 1997, p. 172).

The MBTI community views all people as essentially normal, rather than seeing people as either "normal" or "abnormal." Pearman and Albritton state:

> Since its first publication in 1962, the MBTI is now the most widely used psychological instrument in the world. It has been translated into more than thirty languages, and to date an average of five million people per year take the MBTI in some setting or another. To our knowledge, Jung's model of psychological type, as embodied in the MBTI, is the only theory of human psychology that is based on normal populations and that

> emphasizes the constructive use of differences, rather than simply classifying and defining differences as matters of good-better-best or normal-abnormal outcomes. Jung's notion, honored in Myers and Briggs' work, is that the different styles of perception, judgment, and energy flow are just that—different. One is not inherently better or worse than another. Society may not take kindly to a model in which everybody wins, but it is our contention that this model is the key to successfully navigating the future." (Pearman & Albritton, 1997, p. xiii)

Pearman and Albritton assert that we are all normal but with different expressions of what is normal: "So, who is 'normal'? In large measure, we all are. Our hope is that in these pages you will find the insights into yourself and others that will provide you with the courage to celebrate, in all its many forms, the normalcy of us all" (Pearman & Albritton, 1997, p. xvi).

The HBDI, by Ned Herrmann, "measures a person's preference both for right-brained or left-brained thinking and for conceptual or experiential thinking" (Leonard & Straus, 1997, p. 115). Neither cognitive style is viewed as superior, as both can contribute to the success of a person or organization. In fact if an organization does not have a mix of employees with different styles, and if employees do not respect each other's styles, the result can be quite destructive to the ability of that organization to innovate. As the authors of a recent *Harvard Business Review* article on cognitive styles state: "Preferences are neither inherently good nor inherently bad. They are assets or liabilities depending on the situation. . . . Understanding others' preferences helps people communicate and collaborate" (Leonard & Straus, 1997, p. 113).

Note the stated purpose of both instruments—to improve people and their organizations. Anne Anastasi and Susana Urbina, in their seventh edition of *Psychological Testing,* mentioned another instrument that illustrates this approach to testing:

> There is renewed emphasis on the need for assessment tools that are oriented toward positive mental health rather than psychopathology. . . . The Student Adaptation to College Questionnaire (SACQ-R; Baker & Siryk, 1989) is yet another tool which . . . typifies the application of psychological testing to individual self-understanding and self-enhancement, an application that is a direct outgrowth of the influence of counseling psychology and that is likely to expand greatly in the future. (Anastasi & Urbina, 1997, p. 532)

The established matrix takes a decidedly less positive view of human beings and has a tendency to focus on the abnormal or on people's deficiencies. Carol Tavris in *The Mismeasure of Woman* cuts through the pretense of the established matrix: "Yet when we peer beneath the surface, we find that the old attitude that transforms normal desires and deeds into pathology is alive and well" (Tavris, 1992, p. 177).

There is some empirical evidence that an emphasis on the positive rather than negative aspects of test takers' performance can actually enhance that performance: In a particularly well-designed investigation with seventh-grade students, Bridgeman (cited in Anastasi, 1988) found that "success" feedback was followed by significantly higher performance on a similar test than was "failure" feedback in students who had actually performed equally well to begin with. This type of motivational feedback may operate largely through the goals of the participants set for themselves in subsequent performance and may thus represent another example of the self-fulfilling prophecy (Anastasi, 1988).

View of Outcomes

Test scores historically have served to rank-order people, to establish cutoff points to determine who is "in" or "out," or to place positive or negative labels on people. Answer choices typically have been between a right answer and a wrong answer, or between two answers that result in a person scoring high or low on a scale in which the top or bottom had negative or positive connotations. The emergent matrix tends to report test scores that place the test taker along a continuum, with both ends of the continuum being equally acceptable or positive and with no right or wrong answers—just answers that identify a person's preferences, strengths, and areas for development. In a sense, the shift has been away from scores that yield an up or down ranking and toward ones that indicate right or left poles on a horizontal continuum—with no positive or negative valences assigned to either pole. In fact, the directions to the test taker in the MBTI standard form are: "There are no 'right' or 'wrong' answers to these questions. Your answers will help show you how you look at things and how you like to go about deciding things" (MBTI Form G Self-Scorable, p. 1).

The previously mentioned article in the *Harvard Business Review*, which reviews instruments that measure preferences, draws this key distinction between preferences and traits or abilities: "What we call cognitive differences are varying approaches to perceiving and assimilating data, making decisions, solving problems, and relating to other people. These approaches are preferences (not to be confused with skills or abilities)" (Leonard & Straus, 1997, pp. 112–113). The authors cite the importance of using instruments that are well developed and well validated, and claim that "managers who use instruments with the credibility of the Myers-Briggs Type Indicator (MBTI) or the Herrmann Brain Dominance Instrument (HBDI) find that their employees accept the outcomes of the tests and use them to improve their processes and behaviors. . . . Instruments such as the MBTI and the HBDI will help you understand yourself and will help others understand themselves" (Leonard & Straus, 1997, p. 116).

Locus of Control

The right to privacy and to decide what will be done with one's test results—and who will do it—is equally crucial to the emergent matrix. In the established matrix, testing is conducted by assessment professionals who receive the results and may or may not provide an interpretation of the results to the test taker. In the emergent matrix, the testing is conducted on behalf of the individual, and the results are shared with the individual first. The individual then decides who else will have access to those results and whether he or she wants them shared at all. Anne Anastasi saw this trend back in 1988 when she wrote: "There is growing emphasis, too, on the use of tests to enhance self-understanding and personal development. Within this framework, test scores are part of the information given to the individual as aids to his or her own decision-making processes" (p. 4). . . . "There has been a growing awareness of the right of individuals to have access to the findings in their own test reports" (p. 57).

The ethos of the MBTI community is extremely clear about the centrality of test takers and the primacy of their rights as contrasted with those of the test administrator:

> When presenting type or being introduced to psychological type, it is imperative that the value of the right of self-determination is honored at each juncture of interpretation. By right of self-determination we mean that when you receive the results of the MBTI or any other psychological instrument, you are the expert that interprets it. Your years of feedback from others and reflection on your own behavior take precedence over any other interpretation. . . . Finally, you—the receiver of type-instrument data—should determine your type preferences. . . . You are the final judge. Anyone who says differently should be treated warily. (Pearman & Albritton, 1997, pp. 170–171)

The fact that the vast majority of administrations of the MBTI are given using the self-scoring form helps to ensure that the test taker retains control over who receives their results and what use is made of them.

A second aspect of this issue of centrality of the test taker is a discernible trend toward a focus on the effects of testing or assessment on the test taker, and on designing user-friendly—from the test taker's perspective—tests and assessment practices. For example, *Educational and Psychological Testing: The Test Taker's Outlook* (Nevo & Jager, 1993) examined test takers' perceptions of and attitudes toward psychological tests from a variety of perspectives and with a variety of methods: (a) public-opinion surveys about testing, (b) group interviews about test takers' views, (c) comparison of attitudes toward testing among middle-class and lower-class students based on situational bias, (d) comparison of test takers' views about essay versus multiple-choice exams, (e) the use of examinee feedback questionnaires to elicit their views, (f) comparison between employee selection by use of personal interviews versus by psychometric exams (examinees preferred the psychometric exam, by the way), (g) ideas on how to "humanize" the testing environment and improve the physical conditions of the testing environment, and (h) comparisons of employees' views about psychometric employment tests versus performance appraisals. The editors emphasize the value of focusing on test takers' views: "The contributors to this book share the common professional belief that the examinee's perspective on testing is both important and relevant, and, as such, should be incorporated into the improvement of specific tests and testing in general" (Nevo & Jager, 1993, p. 11).

Tim Rogers argues that testing can no longer be defended from a scientific perspective alone and that tests have a social consequence that cannot be ignored. He states, "Perhaps the most important and emancipating conclusion that can be drawn from the material in this text is that psychological testing is not scientific but is part-and-parcel of the sociopolitical world in which we live. The rhetoric of science is used to promote testing, but at root the enterprise is social and political. The scientific considerations are secondary" (Rogers, 1995, p. 19). This leads to his constructive recommendation to test developers:

> Test development may increasingly begin to reflect the cultural reality experienced by those being tested, rather than revealing the cultural experiences of the test makers. Theoretical constructs may be developed that are integrated into the ongoing cultural context of the group being tested. Tests may be developed to facilitate the manner in which members of a given group can articulate the nature of their problems and

> concerns in their own language, rather than that of the professionals. (Rogers, 1995, p. 792)

He concludes that tests will be judged increasingly by their social impact or consequences: "in the final analysis, it will be the social success of the testing enterprise, not its scientific status, that will dictate the acceptance of the testing movement" (p. 771). "After all, if testing is fundamentally a social activity, then why not evaluate it directly in those same terms? The major bonus of this view is that tests that fulfill this 'new' criterion would have maximal social utility" (p. 788). Social utility to the test taker herself or himself is what Rogers is concerned about, a position that it is diametrically opposed to the focus in the established matrix on institutional and scientific utility.

Conclusion and Recommendations

It is not my intent to suggest that we abandon the established matrix of testing and assessment; rather, I suggest that we augment and enhance that approach with a newer and more promising approach that can blunt many of our critics' arguments and win the public and their representatives over to a favorable perception and stance toward testing. I am not asking that we shift all of our testing and assessment resources to this new type of testing, merely that we devote some of our research and development resources to exploring and experimenting with an approach that has deep roots in our political culture and thus is likely to be embraced instead of excoriated by the American public.

Standardized psychological and educational tests have proven their utility for making many societal decisions, and thus far no more accurate and reliable methods of assessment for decision making have been developed. Anne Anastasi made this claim eloquently and convincingly: "If tests were abolished, the need for making choices, by individuals as well as organizations, would remain. Decision making would have to fall back on such long-familiar alternatives as letters of recommendation, interviews, and grade-point averages. Today these alternative data sources are often used in conjunction with test scores, but not in place of tests. In fact, standardized tests were introduced as one means of compensating for the unreliability, subjectivity, and potential bias of these traditional procedures" (Anastasi, 1988, p. 68). These alternatives to testing have generally proved to be less accurate than tests in predicting school or job performance.

The emergent matrix to test development and use will ensure that testing will remain a thriving enterprise accepted by governmental representatives who have the power to destroy this enterprise we have all built and continue to augment and adapt to changing circumstances.

This emergent approach is founded on the same principles that motivated the American colonies to revolt and establish their independence: a respect for individual self-determination. As an industry and as an applied science, we have too often strayed from our democratic and egalitarian roots, serving instead the interests of the entrenched elite. We need to remind ourselves what led us to become a free country and a safe haven for the millions of oppressed people who left their homelands and came to America to put down new roots on freer soils.

Thanks to decades of continuous development of tests in accord with the established matrix, we have a solid base from which to experiment and innovate along the lines of the emergent matrix. It is my belief that if we succeed with the emergent approach to testing, we will reach and help develop hundreds of millions of people and expand the boundaries of the testing industry far beyond our wildest expectations.

References

Aiken, L. R. (1988). *Psychological testing and assessment* (6th ed.). Newton, MA: Allyn & Bacon.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Baker, R. W., & Siryk, B. (1989). *Student Adaptation to College Questionnaire*. Los Angeles: Western Psychological Services.

Briggs Myers, I. (1977/1999). *Myers-Briggs Type Indicator*. Palo Alto, CA: Consulting Psychologists Press.

Cunningham, G. K. (1986). *Educational and psychological measurement*. New York: Macmillan.

Daves, C. W. (Ed.). (1984). *The uses and misuses of tests*. San Francisco, CA: Jossey-Bass.

Geisinger, K. (Ed.). (1982). *Psychological testing of Hispanics*. Washington, DC: American Psychological Association.

Geisinger, K. & Schafer, W. (Co-chairs). (September 1, 1997). *The rights and responsibilities of test takers*. Joint Committee on Testing Practices Working Group.

Gould, S. J. (1996). *The mismeasure of man.* New York: W. W. Norton and Co.

Hansen, J., Neuman, C. Haverkamp, B., & Lubinski, B. (1997). Comparison of user reaction to two methods of Strong Interest Inventory administration and report feedback. *Measurement & Evaluation in Counseling & Development, 30*(3), 115–127.

Hoffman, B. (1962). *The tyranny of testing.* New York: Crowell-Collier Press.

Kramer, J. J., & Conoley, J. C. (Eds.). (1992). *Hermann Brain Dominance Instrument.* Lake Lure, NC: Applied Creative Services.

Leonard, D., & Straus, S. (1997, July–Aug.). Putting your whole company's brain to work. *Harvard Business Review,* 111–121

Lubinski, D., and Daves, R. (Eds.). (1995). *Assessing individual differences in human behavior.* Palo Alto, CA: Davies-Black Publishing.

Myers, K. (1997). Isabel Briggs Myers' vision—Our work. *Australian Journal of Psychological Type 6*(1), 21–30.

Nevo, B., & Jager, R. S. (Eds.). (1993). *Educational and psychological testing: The test taker's outlook.* Bern, Switzerland: Hogrefe & Huber.

Pearman, R., & Albritton, S. C. (1997). *I'm not crazy, I'm just not you.* Palo Alto, CA: Davies-Black Publishing.

Rogers, Tim B. (1995) *The psychological testing enterprise: An introduction.* Pacific Grove, CA: Brooks-Cole.

Tavris, C. (1992). *The Mismeasure of woman.* New York: Simon and Schuster.

About the Author

Lorin Letendre is chairman and CEO of Consulting Psychologists Press, Inc. He earned his M.A. in political and social sciences from the University of California at Santa Barbara. Among his research interests are issues of assessment for healthy human development and office politics. He is the author of *Understanding American Politics through Fiction*, published by McGraw-Hill. In 1999 he was recipient of both the Arthur Hitchcock Award from the American Counseling Association and the President's Award from the National Career Development Association. He is also a past president of the Association of Test Publishers.

1. This article was originally published in July 1998 in the *Australian Journal of Psychological Type*, *7*, (1–2).
2. Myers-Briggs Type Indicator® instrument is a registered trademark of Consulting Psychologists Press.

Chapter Three

Responding to Testing Needs in the Twenty-First Century With an Old Tool

Lawrence M. Rudner

Abstract

Bayes' theorem is introduced as a method for criterion-referenced testing. This theorem determines the most likely classification for an examinee from a dichotomous choice or through placement on a categorical or interval scale. The application of Bayes' theorem to computer adaptive tests, in which an examinee's ability level is estimated during the testing process and items selected accordingly, is discussed. Relative to item response testing, Bayesian adaptive testing requires less pretesting, needs smaller item pool, can be applied to criterion-referenced and diagnostic testing, can generate classifications based on multiple skills, and requires relatively little statistical knowledge.

Much of modern assessment research and development concentrates on norm-referenced tests, which by definition are designed to rank-order students by placing them on broad continua representing unidimensional traits. The summative information from norm-referenced assessments serves many purposes, but as we enter the twenty-first century, there is a rising call for criterion-referenced information concerning what students know and can do relative to clearly defined desired outcomes of instruction. Although criterion-referenced interpretations of norm-referenced tests are commonplace, the literature from the 1970s and 1980s on criterion-referenced tests can provide some insights to guide current research and practice. As Hambleton and Sireci (1997) point out, the differences between the performance tests of today and the criterion-referenced tests of the 1970s are not fundamental. Both are focused on assessment of what students know and can do.

This article introduces ways of responding to the current clamor for criterion-referenced information using Bayes' theorem—a method that was coupled with criterion-referenced testing in the early 1970s (see Hambleton and Novick, 1973). After introducing Bayes' theorem, I provide some detail demonstrating how it can provide the basis for computer adaptive criterion-referenced tests. I then briefly discuss other potential classroom applications of Bayes' theorem. Specific advantages of using this model are that relatively small data sets are required and that the necessary computations are surprisingly simple.

Bayes' Theorem: A Brief Overview

Rather than placing a student on an ability scale, the goal of a Bayesian approach is to identify the most likely classification for the examinee. This classification may be dichotomous (e.g., master/non-master), polychotomous (e.g., master/at-risk/non-master) or a placement on a categorical or interval scale. A simple example in which the goal is to classify an examinee as being either a master or a non-master is used to illustrate Bayes' theorem. Responses to previously piloted items are used to determine the probabilities of mastery P(M) and non-mastery P(N) and then to classify the examinee based on those probabilities. Lacking any other information about the examinee, let us assume equal prior probabilities, i.e., P(M) = .50 and P(N) = .50. After each item is scored, we will update P(M) and P(N) based on the response to the item.

As givens, we will start with a collection of items for which we have determined the following four probabilities:

1. Probability of a correct response given that the examinee has mastered the material
2. Probability of an incorrect response given that the examinee has mastered the material
3. Probability of a correct response given that the examinee has not mastered the material
4. Probability of an incorrect response given that the examinee has not mastered the material

I will denote these as P(C|M), P(I|M), P(C|N), and P(I|N), respectively. Note that there are different conditional probabilities for each item. These conditional probabilities can be determined from very small-scale, low-cost pilot testing; for example, one approach is to use the percentages of examinees in each group responding correctly or incorrectly. Suppose that on item 1 of the pilot test, 90% of the masters and 40% of the non-masters responded correctly. Because a person responds either correctly or incorrectly, P(C|M) = .90, P(I|M) = .10, P(C|N) = .40, and P(I|N) = .60.

The task then is to update P(M) and P(N) based on the item responses. The process for computing these updated probabilities is referred to as *Bayesian updating, belief updating* (probabilities being a statement of belief), or *evaluating the Bayesian network*. The updated values for P(M) and P(N) are referred to as the *posterior probabilities*. The algorithm for updating comes directly from a theorem published posthumously by Rev. Thomas Bayes in 1763:

$$P(M|C) \times P(C) = P(C|M) \times P(M)$$

Let us suppose our examinee responds correctly to item 1. The probability of a correct response, P(C), is thus 1.0 and by Bayes' theorem, the new probability that the examinee is a master given a correct response is

$$P(M|C) = (.90 \times .5) / 1.0 = .45$$

Similarly, $P(N|C) = P(C|N) \times P(N) = .40 \times .5 = .20$. We can then divide by the sum of these joint probabilities to obtain posterior probabilities, as follows:

$$P'(M) = .45 / (.45 + .20) = .692$$

and

$$P'(N) = .20 / (.45 + .20) = .308.$$

We use these posterior probabilities as the new prior probabilities, score the next item, and again update our estimates for P(M) and P(N) by computing new posterior probabilities. This process continues until all the items have been scored. Equivalently, we could have computed the product of the relevant probabilities (correct or incorrect) for masters and non-masters, then divided by the sum to obtain the last posterior probability.

The Bayesian network defined here is a simple diverging graph. The master/non-master state is causally connected to the set of item responses. When applied to decision-support systems and other expert systems, Bayesian networks are typically much more complex, involving hundreds of interconnected and cross-connected variables (see Lauritzen & Spiegelhalter, 1988; Pearl, 1986). Evaluating such networks is computationally complex. As I have shown here, however, the computations for basic applications are quite simple.

Bayesian Computer Adaptive Testing

Paper-and-pencil tests are typically fixed-item tests in which all examinees answer the same questions within a given test booklet. This is terribly inefficient. Bright individuals have to endure items that cover skills and knowledge they clearly possess. Less able individuals have to suffer through material that is above their ability. These "too easy" and "too difficult" items function like adding constants to an individual's score, providing relatively little if any information about the examinee's true ability. Consequently, large numbers of items and examinees are needed in order to obtain a modest degree of precision, reliability, and validity.

With a computer adaptive test, the examinee's ability level can be iteratively estimated during the testing process, and items can be selected based on a precision-based real-time estimate of the individual's ability. From the pool of items, examinees can be presented with those items that maximize the information about their ability levels. Thus, examinees will receive few items that either are very easy or very hard for them. This tailored item selection results in reduced standard errors and greater precision with only a handful of properly selected items. The time required for testing is greatly reduced, and examinees receive valid, reliable, and legally defensible estimates of their ability. In addition, retesting can occur more frequently without requiring that massive, entirely new item pools be developed and validated.

With the growth of expert systems and the use of artificial intelligence, there has been increasing interest in the use of probability theory and Bayesian networks as a tool to help synthesize observations and generate probabilistic assumptions about current student ability. This information, in turn, may be used to guide the presentation, sequencing, and pacing of instruction. The same mathematical principles have also been proposed as the basis for an attractive form of adaptive testing applicable to a wide range of situations. Relative to item response theory computer adaptive testing (IRT CAT), Bayesian adaptive testing (B-CAT), requires little pretesting and a small item pool. B-CAT can be used with criterion-referenced tests, used to make mastery–non-mastery classifications, incorporated into diagnostic testing, and easily applied to multidimensional assessments. Further, the mathematics of B-CAT are much simpler than those of IRT CAT.

The traditional paradigm for computer adaptive testing is an iterative process with the following steps:

1. A tentative ability estimate is made.
2. All the items that have not yet been administered are evaluated to determine which will be the best one to administer given the current estimate of ability.
3. The best item is administered and the examinee responds.
4. A new ability estimate is computed based on the responses to all of the administered items.
5. Steps 2 through 4 are repeated until a stopping criterion is met.

Bayesian computer adaptive testing follows the same five steps. Instead of estimating ability, however, B-CAT estimates classification probabilities. Frick (1992), and Madigan, Hunt, Levidow, and Donnell (1995) explain how Bayesian networks can be used as the CAT framework. Welch and Frick (1993) provide a excellent and readable overview of the topic. With B-CAT, the goal is to determine the most likely classification for the examinee. This classification may be dichotomous (e.g., master/non-master) or may involve placement on a categorical or interval scale. With B-CAT, conditional probabilities are the givens and posterior probabilities are iteratively estimated. Possible stopping criteria include time, number of items administered, or change in ability estimate. With Bayesian adaptive testing, a desired alpha and beta level can be employed.

To explain B-CAT, I provide an example where the goal is to classify an examinee as being either a master or a non-master. Basically, the new posterior probabilities are computed after each item is administered. One stops administering items when the probability of mastery is sufficiently high or low. Items are selected from the pool of remaining items to maximize information or minimize a loss function.

As givens, let us assume a collection of items for which the four probabilities outlined previously have been determined. We will use the database of four items shown in Table 3.1 (the data for this example come from Welch and Frick, 1993). For the example, we will assume these items are administered sequentially. Ideally, the next item to be administered would be the item that minimizes $P(C_i|M) - P(C_i|N)$; that is, the item most likely to yield the largest change in the posterior probabilities.

Table 3.1. Sample Probabilities of Correct and Incorrect Responses by Masters and Non-masters

	Masters (M)		Non-masters (N)	
Item (i)	$P(C_i|M)$	$P(I_i|M)$	$P(C_i|N)$	$P(I_i|N)$
1	.89	.11	.65	.35
2	.81	.19	.24	.76
3	.92	.08	.47	.53
4	.98	.02	.86	.14

Note that for each *i*, $P(C_i|M) + P(I_i|M) = 1.00$ and $P(C_i|N) + P(I_i|N) = 1.00$. Responses are dichotomous states—an examinee responds either correctly or incorrectly. The goal is to classify the examinee as most likely being a master or a non-master based on his or her responses to selected items. Again, lacking any other information about the examinee, we will assume equal prior probabilities of being a master or non-master (i.e., $P(M) = .50$ and $P(N) = .50$). After each item is given, we will update $P(M)$ and $P(N)$ based on the response to the item.

Let us suppose our examinee responds incorrectly to item 1. By Bayes' theorem, the new probability that the examinee is a master given an incorrect response is

$$P(M|I_i) = P(I_i|M) \cdot P(M) / P(I_i)$$

We know that the examinee has responded incorrectly, so $P(I_i) = 1.00$ and $P(M|I_i) = .11 \times .5 = .055$. Similarly, $P(N|I_i) = P(I_i|N) \times P(N) = .35 \times .5 = .175$. We can then divide by the sum of these joint probabilities to obtain posterior probabilities, as follows:

$$P'(M) = .055 / (.055 + .175) = .239$$

and

$$P'(N) = .175 / (.055 + .175) = .761$$

We next use these posterior probabilities as the new prior probabilities, select a new item, and again update our estimates for P(M) and P(N) by computing new posterior probabilities. We iterate the process until some specified stopping criterion is reached. Wald's (1947) Sequential Probability Ratio Test appears to be favored in the literature.

To continue the example, let us assume that the examinee responds correctly to item 2, incorrectly to item 3, and incorrectly to item 4. Table 3.2 shows the resultant probabilities for all four items.

Table 3.2. Calculations for Probability of Mastery Based on Four Sample Responses

Item (i)	Response (R_i)	State(S)	Prior Probability	$P(S\|R_i)$	Joint Probability	Posterior Probability
I	I	Master	.500	.11	.055	.239
		Non-master	.500	.35	.175	.761
2	C	Master	.239	.81	.194	.515
		Non-master	.761	.24	.183	.485
3	I	Master	.515	.08	.041	.138
		Non-master	.485	.53	.257	.862
4	I	Master	.138	.02	.003	.024
		Non-master	.862	.14	.121	.976

At each iteration, the subsequent item can be selected to maximize the expected change in the posterior probability. After administering these four items, the probability that our examinee is a non-master given this response pattern is .976. Had we set a minimum posterior probability of .975 ($1 - \alpha/2$) as the stopping rule, we could then terminate item administration.

In theory, this approach to CAT has the advantages of IRT CAT plus several crucial advantages of its own:

- It can incorporate a small item pool.
- It is simple to implement.
- It requires little pretesting.
- It can be applied to criterion-referenced tests.
- It can be used in diagnostic testing.
- It can be adapted to yield classifications on multiple skills.
- It is easy to explain to non-statisticians.

In recent years there has been growing theoretical interest in B-CAT among the educational testing community (De Ayala, 1990; Frick, 1992; Lewis & Sheehan, 1990; Segall, 1996; Spray & Reckase, 1996; van der Linden & Hambleton, 1997). There have also been a handful of small studies evaluating B-CAT. De Ayala (1990), Jones (1993), Spray and Reckase (1996), and Welch and Frick (1993) all found advantages to B-CAT relative to other forms of adaptive testing. These studies, however, were typically limited to one examination and to relatively small samples. B-CAT is also featured as the engine behind at least one large company offering intelligent tutoring system development services (Gemini Learning Systems Inc.: http://www.gemini.com).

Classroom Applications

The basic framework described in this article is applicable to a wide range of settings. For example, the framework can be used to score a diagnostic pretest. Here the pretest would cover a variety of skills. A pilot test would determine the probabilities of responding correctly for people who have mastered each skill and the probabilities for those that have not done so. After the test is given to an individual, the probabilities of mastery for each skill could be computed. The resultant list would identify which skills have been mastered and which are likely in need of attention. One could go further and model specific misconceptions (e.g., the examinee sums denominators when adding fractions). Here the relevant probability would be likelihood of selecting a particular incorrect option (or generating a particular type of wrong answer) given that an examinee has a specific misconception. Such a test would not only provide mastery information but identify specific areas to correct.

The framework is also applicable to multidimensional items and tests. One could write items, for example, that require the application of mathematical skills to solve a science problem. A pilot test would need to be administered to compute the probability of responding correctly to each item given mastery of the mathematics skills and the probability of responding correctly given mastery of the science skills. The single test with complex items could then be scored, using the Bayes' theorem and information about each skill area.

Finally, the framework can be embedded in an intelligent tutoring system to determine mastery after each instructional unit, tailor individualized instruction to characteristics of the student, and adapt that instruction as the student learns material. This would again require a collection of pretested items that assess the concepts covered by each instructional unit.

Research Questions

Some concern has been raised concerning the sensitivity of Bayesian networks to misspecified prior probabilities. This is not really a concern with B-CAT, as the system will converge after the administration of only a few items, as it does with IRT CAT. Our concerns are (a) whether B-CAT truly leads to efficient and accurate state classifications, and (b) the sensitivity of Bayesian networks to misspecifications of the conditional probabilities. Probability theory defines expectations over large data sets and large samples. Yet, with B-CAT, we are interested in making inferences about individuals based on small data sets. Thus, B-CAT is a theory that has yet to be demonstrated to work in realistic situations. Bayesian conditional probabilities are based on either qualitative judgments or sampled empirical data. In either case, the specified conditional probabilities are not the same as true conditional probabilities. One is working with estimates, not true values, and the resulting inherent error can seriously bias the results. Shrinkage could be an issue, and the effect that error in the conditional probabilities may have on the posterior probabilities and the number of items needed is not clear.

References and Resources

One can easily experiment with simple Bayesian networks using any of a large variety of readily available, free software packages. A search on the Internet in November 2000 for "Bayesian Network

Software Packages" yielded more than 20 free packages that could potentially be applied. Two that I have tried are Hugin Lite and Genie.

Bayes, T. (1763). Essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London,* 53, 370–418.

Chamiak, E. (1991). Bayesian networks without tears. *AI Magazine,* Winter.

De Ayala, R. J. (1990). A simulation and comparison of flexilevel and Bayesian computerized adaptive testing. *Journal of Educational Measurement,* 27(3), 227–239.

Frick, T. W. (1992). Computerized adaptive mastery tests as expert systems. *Journal of Educational Computing Research 8*(2), 187–213.

Hambleton, R. K., & Novick, M. R. (1973). Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement, 10*(3), 159–170.

Hambleton, R. K., & Sireci, S. G. (1997). Future directions for norm-referenced and criterion-referenced achievement testing. *International Journal of Educational Research, 27*(5), 379–393.

Jones, W. P. (1993). Real-data simulation of computerized adaptive Bayesian scaling. *Measurement and Evaluation in Counseling and Development, 26*(2), 143–151.

Lauritzen, S. L., & D. J. Spiegelhalter. (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of the Royal Statistical Society,* Series B, *50*(2), 157–224.

Lewis, C., & Sheehan, K. (1990). Using Bayesian decision theory to design a computerized mastery test. *Applied Psychological Measurement, 14*(4), 367–386.

Madigan, D., Hunt, E., Levidow, B., & Donnell, D. (1995). *Bayesian graphical modeling for intelligent tutoring systems.* (Technical Report). Seattle: University of Washington.

Pearl, J. (1986). Fusion, propagation, and structuring in belief networks *Artificial Intelligence, 29,* 241–288.

Segall, D. O. (1996). Multidimensional adaptive testing. *Psychometrika, 6*(2) 331–54.

Spray, J. A., & Reckase, M. D. (1996). Comparison of SPRT and sequential Bayes procedures for classifying examinees into two categories using a computerized test. *Journal of Educational and Behavioral Statistics, 21*(4), 405–414.

van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory.* London: Springer Verlag.

Wald, A. (1947). *Sequential analysis.* John Wiley and Associates.

Welch, R. E., & Frick, T. (1993). Computerized adaptive testing in instructional settings. *Educational Training Research and Development, 41*(3), 47–62.

About the Author

Lawrence M. Rudner is director of the ERIC Clearinghouse on Assessment and Evaluation and assistant director of the Maryland Assessment Research Center for Education Success at the University of Maryland, College Park. He earned his Ph.D. in educational psychology and evaluation from Catholic University of America and also holds an MBA in finance from the University of Maryland, College Park. Among his areas of interest are educational measurement and information science. Co-author of *What Teachers Need to Know about Assessment* (Washington, DC: National Education Association), Rudner also serves as editor of *Practical Assessment, Research and Evaluation* and associate editor of *Applied Measurement in Education.* In addition, he chairs the AERA Telecommunication Committee.

Counselor Education

Assessment in Counselor Education: Admissions, Retention, and Capstone Experiences
Irene Mass Ametrano and Sue A. Stickel

Revitalizing the Assessment Course in the Counseling Curriculum
Albert B. Hood

The Pedagogical Basis for Multifaceted Assessment in Counselor Education
Barbara D. Yunker and Mary E. Stinson

Chapter Four

Assessment in Counselor Education: Admissions, Retention, and Capstone Experiences

Irene M. Ametrano
Sue A. Stickell

Abstract

The development of effective procedures for assessing the competency of counselors-in-training is one of the greatest challenges currently facing counselor educators. The responsibility for assessment is documented in codes of ethics and standards of all counseling professional organizations. Students should be assessed at three stages—at admission, during the program, and on graduation—on both academic competency and personal issues that might interfere with professional ability. The assessment procedures used at one program are described to illustrate how such assessment might be accomplished.

Counselor educators are increasingly serving as gatekeepers for the counseling profession. As part of this role, they grapple with how to assess counselor trainees' potential to be effective counselors. Separate assessment questions arise at the point of students' admission to an educational program, as students progress through the training program, and at exit or graduation. Making admissions decisions involves determining criteria for who has the potential to become a counselor. Decisions regarding retention in the program require an ongoing assessment of how or whether the necessary competencies are developing. At graduation, counselor educators in most states make recommendations for licensure and must determine whether or not the student has acquired the requisite competencies. This paper will review relevant literature addressing these questions and outline the assessment procedures used in one program.

Codes of ethics, accreditation standards, and recent legal cases provide a foundation for the assessment responsibilities of counselor educators. The American Counseling Association *Code of Ethics and Standards of Practice* (1995) addresses evaluation, limitations, and endorsement of students and supervisees:

> Counselors clearly state to students and supervisees, in advance of training, the levels of competency expected, appraisal methods, and timing of evaluations for both didactic and experiential components. Counselors provide students and supervisees with periodic performance appraisal and evaluation feedback throughout the training program. (Section F.2.C) . . . Counselors, through ongoing evaluation and appraisal, are aware of the academic and personal limitations of students and supervisees that might impede performance. Counselors assist students and supervisees in securing remedial assistance when needed, and dismiss from the training program supervisees who are unable to provide competent service due to academic or personal limitations. (Section F.3.A) . . . Counselors do not endorse students or supervisees for certification, licensure, employment or completion of an academic degree training program if they believe students or supervisees are not qualified for the endorsement. (Section F.1.H.)

The Association for Counselor Education and Supervision (ACES) identifies similar assessment responsibilities of supervisors in its *Ethical Guidelines for Counselor Educators and Supervisors* (1993), which states, "Supervisors have the responsibility of recommending remedial assistance to the supervisee and of screening from the training program, applied counseling setting, or state licensure those supervisees who are unable to provide competent professional services" (Section 2.12).

Finally, the Council for the Accreditation of Counseling and Related Educational Programs (CACREP, 1994) has established standards that require counselor-education programs to have clear admissions criteria, as well as selection and retention procedures:

> When evaluations indicate a student's inappropriateness for the program, faculty assist in facilitating the student's transition out of the program and, if possible, into a more appropriate area of study. (Section F.2.C) . . . Admissions criteria, as well as selection and retention procedures, should consider qualities such as the applicant's potential success in forming interpersonal relationships; aptitude for graduate level study; and openness to self-examination and personal and professional self-development. (Section V.K.)

The responsibility of the counselor-education program for ensuring the competence of its graduates is illustrated in a recent lawsuit against Louisiana Technical University. A graduate of the counseling program was sued by a client for allegedly encouraging a dual relationship. The client also sued the university for failure to sufficiently train the counselor (Custer, 1994).

Admissions

Admissions decisions historically have been based on academic and other traditional predictors including undergraduate grade-point average (GPA), Graduate Record Examination scores, and letters of recommendation (Bradey & Post, 1991; Gimmestad & Goldsmith, 1973; Hosford, Johnson, & Atkinson, 1984). Bradey and Post (1991) found little data to support academic criteria as predictors of counselor competency and recommended developing effective ways to evaluate criteria such as interpersonal competence, openness to professional self-development, and openness to the values and opinions of others. Interviews or observation of applicant interactions, or both, would facilitate this type of assessment. Hayes (1997) noted a lack of clear-cut guidelines for choosing the most appropriate and effective screening methods. Procedures tend to vary widely from program to program.

Assessing the applicant's/student's mental state, or emotional problems that may prevent the person from working effectively with clients, is necessary. The notion of the "wounded healer" (Maeder, 1989), that people with psychological problems are drawn to the helping professions, is controversial and the data are not consistent. We do know, however, that in order for counselors to be effective with clients, their own problems cannot interfere. The counselor's first responsibility is to do no harm to the client. White and Franzoni (1990) found that on six of seven Minnesota Multiphasic Personality Inventory (MMPI–2; Butcher et al., 1989) scales, counselors-in-training had higher levels of psychological disturbance (depression, hysteria, psychological deviance, paranoia, psychasthenia, schizophrenia) than the general population. There was no difference in social interest, locus of control, and coping.

How do counselor educators identify those applicants whose psychological state is likely to interfere with their providing competent services to clients? More thorough screening at admissions should reduce the number of students who must be dismissed once they are in the program. Hayes (1997) found little evidence in the literature that counseling programs are using standardized instruments to assess mental disorders in applicants. Increasingly program representatives are using interviews in an informal way, but they are generally not using standardized or even systematic assessment methods. They look for characteristics such as "active mental disorder," "evidence of pathology," "awareness of influence on others," capable/appropriate interpersonal skills, "understanding of self," "inappropriate behaviors," and so on (Hayes, 1997). Without a standardized instrument, however, this is a more subjective process than assessing academic performance.

Specific information is needed concerning the personal characteristics that have been shown to limit a counselor's effectiveness. How and when we assess these is another issue. Hayes (1997) gives the example of a program that requires applicants to take an introductory course that includes small-group work. Students are rated on a scale of 1 to 5 on 13 characteristics including open-mindedness, tolerance of ambiguity, objectivity, sense of humor, willingness to learn and grow psychologically, emotional stability, personal security, and confidence. After an extensive literature review, Frame and Stevens-Smith, (1995) identified nine personal characteristics that are necessary for counselor development: being open, flexible, positive, cooperative, willing to use and accept feedback, aware of impact on others, able to deal with conflict, able to accept personal responsibility, and able to express feelings openly and appropriately. Students in the program are evaluated on these at the midpoint and the end of every course. Baldo, Softas-Nall, and Shaw (1997) defined substandard behaviors, including failure to demonstrate empathic capacity, maturity of judgment, ability to work closely with others, capacity to handle stress, and tolerance for deviance.

Several broad characteristics emerge after reviewing these studies: (a) openness to self-examination (willingness to use and accept feedback, awareness of impact on others, willingness to accept personal responsibility, willingness to learn and grow psychologically); (b) potential for effective interpersonal relationships (awareness of impact on others, ability to work closely with others, empathic capacity, ability to deal with conflict, open and appropriate expression of feelings); (c) open-mindedness (tolerance for deviance and for ambiguity); and (d) emotional stability (capacity to handle stress).

Retention and Dismissal

No matter how good admissions procedures are, some students who cannot meet academic standards or whose personal problems and characteristics interfere with their effectiveness will be admitted to counselor-education programs. Olkin and Gaughen (1991) found that counselor educators often identify problem students through supervised clinical experiences. Problems include poor clinical skills; interpersonal problems; refusal to accept constructive feedback or directions; and intrapersonal problems such as substance use, personality disorders, and immaturity.

Baldo, Softas-Nall, and Shaw (1997) describe a process for review of students' progress in the program and processes for remediation, voluntary resignation, and dismissal from the program. They stress the importance of (a) documentation so that faculty judgments are not seen as capricious or prejudicial and (b) dismissal decisions being made by the entire faculty. Other procedures that insure the student's due process include (a) the student and faculty member are informed of problem areas and methods of remediation; (b) a written plan for remediation is approved by the faculty and signed by the student; (c) the student has the opportunity to present his or her case to the faculty; (d) and an appeals procedure is available. Frame and Stevens-Smith (1995) describe a process that involves the development of a policy statement expressing the faculty's belief in the "essential function" of personal characteristics in the development of ethical and competent counselors. This statement, along with the Personal Characteristics Evaluation Form, is published in the student handbook. Students are required to read the handbook and sign a statement that they have read and will abide by the policies. All syllabi include a statement about professional characteristics and their regular evaluation. Clear steps to follow when problems are identified have been identified, and remediation opportunities are offered if seen as appropriate.

Exit or Graduation

Recently, faculty in counselor-education programs have begun to re-examine their final evaluation methods (Carney, Cobia, & Shannon, 1996). The assessment of a student's ability to apply acquired knowledge and appropriateness for the profession cannot be accomplished by traditional methods such as comprehensive examinations or theses. The portfolio is one way of assessing multiple dimensions that make up counseling effectiveness, however, particularly if the portfolio is used as an adjunct to other methods.

Portfolios have been used in two ways: to document a student's progress over time (developmental or formative evaluation), and to show a student's best work (summative evaluation). It is possible to use portfolios for both formative and summative purposes. In counselor education, portfolios have been used primarily as opportunities for self-reflection or self-assessment by the student. Reviewing portfolios periodically with the student allows for remediation. In this way, a portfolio could be integral to an ongoing evaluation process (Baltimore, Hickson, George, & Crutchfield, 1996).

Using portfolios to demonstrate a student's best work has been discussed less frequently in the counselor-education literature. Carney and colleagues (1996) recommend that such assessment focus on criteria including ability for self-reflection; counseling skills; application of knowledge; professional identification; and ability in specialty areas such as community or school counseling. Contents of portfolios would include research papers, treatment plans, audio and video tapes of practice, progress notes, self-evaluations, and other items. The challenge for faculty is to develop criteria to evaluate each of these components.

Program Example

In order to select students who have both the academic potential to succeed in graduate school and the personal characteristics to be effective counselors, in the Department of Leadership and Counseling at Eastern Michigan University we have developed an extensive two-phase admissions process. By doing a more thorough assessment at the point of admission, we hope to minimize the need for dismissal once students have begun the program. Our admissions screening considers multiple variables, including aptitude for graduate study, career goals, writing ability, and potential for effectiveness as a counselor. In the first phase of the process, faculty members assess the applicant's aptitude for graduate study by considering undergraduate GPA or the GPA from another graduate degree, which must be at least 2.75 for an undergraduate, or 3.3 for a graduate, degree. Although all applicants must take the Graduate Record Examination for admission to graduate programs in the College of Education, we do not consider these scores unless the applicant does not meet the minimum GPA requirement. The applicant's letter of intent is used to assess the extent to which his or her career goals match program goals, as well as writing ability, defined as clarity of expression, organization, and grammar. In this phase of the process, potential for effectiveness as a counselor is assessed by reviewing the applicant's resume and letters of recommendation. The resume of an applicant who has seriously thought about counseling as a career would reflect involvement in personal and professional growth activities and a variety of life and professional experiences. Faculty reviewers rate the letter of intent, resume, and letters of recommendation on a five-point Likert scale from exceptional to unacceptable. Based on these ratings, an applicant may be invited for an interview or screened out, or the application held for discussion with other faculty members.

In the second phase of the process, selected applicants come to campus to participate in group and individual interviews. Assessment during these interviews focuses on the applicant's personal characteristics and potential for success as a counselor. In the group interview, applicants are assigned to a small group, which is given a task to complete. Faculty observe the group interaction and rate each applicant on behaviors considered to be facilitative in interpersonal interactions using a five-point Likert scale. Examples of these behaviors include willingness to listen to others, attempts to understand others, acceptance of difference, openness, and appropriateness of contributions. The purpose of this activity is to identify applicants whose behavior is not facilitative and who, therefore, may be ineffective in a counseling relationship. In the individual interview, each applicant meets with a faculty member and responds to three questions that focus on the applicant's career goals and decision to apply to this graduate program, self-perceptions about areas of strength and weakness, and experiences with people who are different. The interviewer rates the applicant's response to each question on a five-point Likert scale and, based on these ratings, makes a recommendation regarding admission. The faculty then meets to discuss each applicant's ratings from the group and individual interviews, and final admission decisions are made.

Our portfolio process is in a much earlier stage of development than is the admissions process. The portfolio can best be described as a formative assessment and is presented to students as an opportunity to present a collection of evidence of their knowledge, accomplishments, and growth during the program. Contents are to reflect several areas including the student as a new professional (statement of goals and philosophy, resume, professional disclosure statement, etc.); professional and personal growth and development (memberships, presentations, conference attendance, recognition/awards, volunteer experiences, etc.); academic growth and development (assessment profile, group plan, research proposal, case presentation, etc.); and counseling skills and experience (rating forms from skills classes, clinical internship evaluations, skill demonstration on video, treatment plans, etc.).

Once a year, a portfolio symposium is held in a format similar to a conference poster session. Students display their portfolios and discuss them with other students, faculty, administrators, and community members. Although faculty members do provide each student with written feedback about the portfolio, specific criteria for assessment have not been developed at this point.

The development of effective assessment procedures for making admissions decisions, for use as students progress through programs, and for determining which students graduate and become credentialed to provide counseling services is clearly among the greatest challenges currently facing counselor educators. It is imperative that research and dialogue continue to address these issues.

References

American Counseling Association. (1995). *Code of ethics and standards of practice*. Alexandria, VA: Author.

Association for Counselor Education and Supervision. (1993). *Ethical guidelines for counselor educators and supervisors*. Alexandria, VA: Author.

Baldo, T. D., Softas-Nall, B. C., & Shaw, S. F. (1997). Student review and retention in counselor education: An alternative to Frame and Stevens-Smith. *Counselor Education and Supervision, 36*, 245–253.

Baltimore, M. L., Hickson, J., George, J. D., & Crutchfield, L. B. (1996). Portfolio assessment: A model for counselor education. *Counselor Education and Supervision, 19*, 197–206.

Bradey, J., & Post, P. (1991). Impaired students: Do we eliminate them from counselor education programs? *Counselor Education and Supervision, 31*, 100–108.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., Hathaway, S. R., & McKinlay, J. C. (1989). *Minnesota Multiphasic Personality Inventory–2*. Minneapolis, MN: University of Minnesota Press.

CACREP. (1994). *Accreditation standards and procedures manual*. Alexandria, VA: Author.

Carney, J. S., Cobia, D. C., & Shannon, D. M. (1996). Clinical and comprehensive evaluation of counselors-in-training. *Counselor Education and Supervision, 36*, 122–132.

Custer, G. (1994, November). Can universities be liable for incompetent grads? *APA Monitor, 25*(11), 7.

Frame, M., & Stevens-Smith, P. (1995). Out of harm's way: Enhancing monitoring and dismissal processes in counselor education programs. *Counselor Education and Supervision, 35*, 118–129.

Gimmestad, M. J., & Goldsmith, E. B. (1973). Admissions policies for graduate programs in counselor education. *Counselor Education and Supervision, 12*, 172–177.

Hayes, L. (1997, May). Today's grad schools are using stringent standards to screen applicants. *Counseling Today, 40*, 1, 6.

Hosford, R. E., Johnson, M. E., & Atkinson, D. R. (1984). Academic criteria, experiential background, and personal interviews as predictors of success in a counselor education program. *Counselor Education and Supervision, 23*, 268–275.

Maeder, T. (1989). Wounded healers. *Atlantic Monthly,* January, 37–47.

Markert, L. F., & Mohke, R. H. (1990). Changes in counselor education admissions criteria. *Counselor Education and Supervision, 30*, 48–57.

Olkin, R., & Gaughen, S. (1991). Evaluation and dismissal of students in master's level clinical programs: Legal parameters and survey results. *Counselor Education and Supervision, 30*, 276–288.

White, P. E., & Franzoni, J. B. (1990). A multidimensional analysis of the mental health of graduate counselors in training. *Counselor Education and Supervision, 29*, 258–265.

About the Authors

Irene M. Ametrano is professor and counseling program coordinator in the Department of Leadership and Counseling, Eastern Michigan University, Ypsilanti, MI. She earned her Ed.D. in counseling psychology from Rutgers University. Among her research interests are crisis intervention and ethical and legal issues in counseling. Recent publications include the chapter "Preventing School Violence: Training for School Counselors," with S. A. Stickel and Y. C. Callaway, appearing in Sandhu, D. S. & Aspy, C. B. (Eds.) *Violence in American Schools: A Practical Guide for Counselors* (Alexandria, VA: ACA, 2000). Ametrano also serves on the editorial board of *Counselor Education and Supervision.*

Sue A. Stickel is a professor in the Department of Leadership and Counseling, Eastern Michigan University, Ypsilanti, MI. She earned her Ph.D. in counselor education at the University of Wyoming. Her publications and research focus on school counseling and school reform. Recent publications include the chapter "Preventing School Violence: Training for School Counselors," with I. M. Ametrano and Y. C. Callaway, appearing in Sandhu, D. S. & Aspy, C. B. (Eds.) *Violence in American Schools: A Practical Guide for Counselors* (Alexandria, VA: ACA, 2000).

Chapter Five

Revitalizing the Assessment Course in the Counseling Curriculum

Albert B. Hood

Abstract

Although some have claimed that testing is no longer an important function of counselors, studies indicate that counselors in most settings interpret results of psychological assessment instruments. In educational settings, the counselor is the professional most likely to have training in the interpretation of tests, and teachers and administrators frequently rely on the counselor to do so. The assessment course in the graduate curriculum continues to be necessary and must equip trainees for this function while stressing the counseling aspects of this role.

Psychological testing has long been regarded as an important function for the counselor. The psychometric tradition from which psychological testing emerged is considered a core foundation of counseling. In current practice counselors use a variety of assessment instruments ranging from intellectual to vocational to personality measures. Testing is an efficient method of getting accurate information and conveying it, through test interpretations, to the client. Counselors frequently use tests as a means of getting to know and understand their clients' personality, vocational interests, intelligence, or aptitudes. When good tests are used—those that have appropriate reliability and validity for the task at hand—the counselor, through the wise use of test integration, is able to achieve insight into the client more rapidly than is possible through an interview alone. Testing can have a powerful impact on clients, because there is a mystique to testing that enhances the message that test interpretation gives to them. Many clients are able to look at themselves more realistically if they are told about their strengths and problems during a test interpretation than if they simply hear the information from their counselor. Psychological testing can clarify a client's personality or vocational interests and abilities, thereby accelerating counseling and saving the client money and valuable time.

For all of these reasons, an assessment course is required in all master's-level counselor training programs accredited by the Council for the Accreditation of Counseling and Related Educational Programs. The view has been expressed, however, that this type of course is no longer relevant as counselors are moving away from assessment and becoming more involved with activities involving personal and group counseling and psychotherapy (Bradley, 1994; Goldman, 1994b). The results of studies of the use of tests in counseling contradict this assertion, however. In an early-1980s survey of a large number of college counseling agencies, Zytowski and Warman (1982) found that more than 95% made use of psychological tests. Only nine of the total of 198 agencies did not. Specifically, 92% reported using the Strong, 81% the WAIS, 72% the Minnesota Multiphasic Personality Inventory (MMPI–2; Butcher et al., 1989), 80% the Edwards, 67% the Bender-Gestalt, and 65% the DAT. More recently, Watkins, Campbell, and Nieberding (1994) in a survey of more than 600 counselors from a variety of different types of agencies found that 81% reported using assessment instruments. The MMPI, the Strong and the WAIS-R were the assessment instruments most often employed by counselors in community-based settings (Bubenzor, Zimpfer, & Mahrle, 1990). Recently, Elmore, Ekstrom, and Schafer (1998) reported that 91% of school counselors often or occasionally interpreted test scores to students; 82% did so to their parents; and 81% to teachers, administrators, and other professionals. In a survey of more than 400 members of the American School Counseling Association (ASCA), respondents reported that they spent one to five hours or more per week working with tests and 67% believed that testing was an important part of their work (Elmore, Ekstrom, Diamond, & Whittaker, 1993). In his review of various surveys of the use of tests in counseling agencies, Watkins (1991) commented on the remarkable stability of test use over three decades. He concluded that (a) psychological assessment is a major component of counselor-training programs; (b) that most practicing counselors, regardless of work setting, provide assessment services and spend a fair portion of their professional time doing so; and (c) the types of assessment instruments that counselors use are very diverse.

Several decades ago Goldman (1972, 1994a) suggested that the "marriage" between testing and counseling had failed. In fact, it is obvious from the results of these surveys that, as Prediger (1994) stated, testing and counseling is a "marriage that has prevailed"; although the initial honeymoon may be over, the marriage has been sustaining and mutually embracing, and continues to be so having now passed its golden anniversary. Thus a psychological assessment course must continue to be required in the curriculum of a graduate-level counselor-education program.

Important Themes for the Assessment Course

In any appraisal course for counselors, it should be stressed that in a counseling setting tests are used differently than in other settings where tests are employed to make decisions about individuals. In counseling tests are designed to be used by the clients themselves and only in the ways the clients decide. The course should emphasize the integrated manner in which testing is approached in counseling (Duckworth, 1990), namely the following:

Testing as an aid to client problem solving: For the individual who is coming in for help with either personal or career issues, the objective picture that a test can give is an invaluable aid in problem solving. Tests can provide clients with a way of stepping back to view their abilities, emotions, and interests in an objective, non-emotional manner—to see themselves from a different perspective and see how they compare to other people. Because counselors look for a client's strengths as well as weaknesses and problems, they use tests that report on aspects of the normal personality as well as those that report what is abnormal (Duckworth, 1990). Tests such as the Myers-Briggs, the California Psychological Inventory, and more recently the NEO have become popular, indicating the increasing use of tests designed to measure normal behavior. This approach to testing enlists the power of the client as well as the expertise of the counselor to effect instructive change.

Testing as an aid in decision making: The use of testing as an aid in decision making begins by giving clients a say in the tests that are to be used. Two of the advantages of doing so are that clients will have greater motivation to answer inventories accurately and will be more likely to use the test results to make personal changes. In order for clients to use the information for change, it is imperative that the test results be given in a language that they can understand. In addition, counselors must use non-pejorative language that describes the test information without attaching a value judgment to it. Doing so avoids the client defensiveness that typically arises when value-laden terms or psychological jargon are used.

Testing as a psychoeducational tool: Tests themselves are considered psychoeducational tools, and counselors employ them in an educative, facilitative style to enhance client exploration and reflection. In this way tests are one method by which counselors can engage clients in a psychological, educational experience. Test results provide clients with important information about themselves that can be of value for personal growth purposes. Therefore, counselors actively involve clients in the assessment process and act as both teacher and facilitator during the experience (Duckworth, 1990). For example, career assessment usually includes the components of an individual's interests, values, skills, and abilities. These may be assessed in an integrated fashion through computer-based guidance systems or through the administration of different types of assessment instruments, participation in simulations and exercises, or the use of various performance measures. Through a factor analysis, Swanson (1993) showed clearly that interest, abilities, and skills are sufficiently distinct to be considered separate constructs worthy of independent assessment. One of the problems of using self-estimates rather than objective measures relates to their accuracy: the relationship between self-evaluation and actual ability is often very low, with mean correlations generally running below .3 (Lowman & Williams, 1987; Mabe & West, 1982). Self-ratings typically rank abilities significantly higher than do objective measures. Moreover, self-evaluations of skills and abilities show little relationship to college admissions test scores (Swanson & Lease, 1990) and contribute less to predictions of occupational attainment 11 years after high school than do ability assessments (Austin & Hanisch, 1990). Thus standardized tests and inventories contribute important information for clients to consider.

Counselors' Responsibilities

Counselors also must be able to understand and interpret the results of various tests that are not administered specifically for use in the counseling process. Tests are increasingly being employed in attempts to solve educational and social problems. National testing programs, ability-to-benefit laws, required assessments in job-training programs; credentialing tests; and mandatory state and local testing programs affect the lives of counselors daily. Tests may be given for assessing student achievement, for college admissions, for personnel selection, or for clinical diagnosis, and these results may or may not be used in counseling or therapy sessions. To adequately understand and interpret such test results—which have varying reliabilities and validities and usually are reported in terms of standard scores, T-scores, or percentiles on varying norm groups—counselors must have an understanding both of the tests themselves and of important measurement concepts. Therefore, this information must be taught in any appraisal course.

Goldman (1992) has suggested that because counselors are not qualified to understand fully all of the psychometric qualities of psychological tests, they should not attempt to administer or interpret them, or even be trained in their use. The fact is, however, that a variety of standardized tests are used in most settings in which counselors work, and they are not only expected to be knowledgeable about these tests, but in many instances are by far the most qualified personnel to interpret them (Tennyson, Miller, Skovolt & Williams, 1989). This is especially true in educational settings, where paper-and-pencil IQ tests and basic skills tests are used in elementary schools; and achievement tests, academic aptitude tests, multiple aptitude tests, and interest inventories are used in high schools and colleges (Engen, Lamb, & Prediger, 1982). Elmore, Ekstrom, Diamond, and Whittaker (1993) reported that teachers see test interpretation as an important part of the counselor's role and expect to be able to turn to the counselor for help with testing questions. An appraisal course must prepare counselors to interpret the results of these tests to teachers, to administrators, to parents, and to the students themselves.

Graduate training programs for school administrators and teachers typically do not require instruction in measurement (Hills, 1991), and in the programs that do, these professionals often find the course content irrelevant (Impara & Plake, 1995). Of the professionals in the school setting, only counselors and school psychologists are likely to have been exposed to formal coursework in assessment. Administrators, in fact, often receive less training in basic assessment than the teachers whose work they are supposed to supervise (Stiggins, 1991). Although teachers use assessment extensively in the classroom, the training they receive is often inadequate to the task. There have also been advances in the field of testing that have rendered out of date the skills of individuals without recent training in tests and measurement. Considerable research indicates substantial deficiencies in teachers' knowledge of assessment practices. There are only four states that require such courses for prospective teachers (Hills, 1991).

If teachers and administrators are ill prepared to conduct the assessment tasks undertaken in the classroom, who then is able to do so? Who can serve as a resource for the teacher when he or she has an assessment problem? A survey showed that both administrators and teachers often report that they rely on counselors to provide answers to questions about testing (Impara & Plake, 1995). School counselors need to be trained to understand and interpret the various scores that might be found on a typical standardized test report, such as equivalents or percentile bands or stanines. Counselors have a basic understanding of the concept of reliability and errors of measurement, whereas many teachers and administrators do not. Being able to explain this concept to others and to use that notion to help teachers understand and interpret test scores and how they can be used in the assignment of grades can be very helpful. Although this application is typically considered beyond the scope of the counselor's role in the school, nevertheless teachers often ask counselors for assistance in this and other areas related to testing (Impara & Plake, 1995). Until the responsibility for expertise in testing is shifted to some other professional, school counselors will continue to be regarded as the "experts" in the school in the area of testing. This carries a certain level of responsibility for being able to correctly interpret and explain the scores from standardized tests.

If counselors are not qualified to interpret and explain tests, then should this responsibility be left up to much less-qualified classroom teachers or principals? Who would prefer that the ASVAB be interpreted to students by a recruiting sergeant instead of a counselor? The counselor is the professional who has had a solid course in psychological testing and training in interviewing, counseling skills, and working with persons as individuals; therefore, the counselor should be the professional with this responsibility. Who else in the typical high school has an understanding of a PSAT score of 70, an ACT score of 22, or an SAT Verbal score of 325 and the meaning and potential impact of such large differences?

Certainly counselors in mental-health settings should not be expected to be able to develop an MMPI interpretation worthy of a Jane Duckworth or a John Graham; still, they should understand the difference on the MMPI between a T-score of 50 (normal) and one of 75 (severe) on the Depression scale, even if they are dependent upon others for in-depth clinical diagnoses and evaluations. If most counselors do not have the necessary knowledge and skills to use tests, it is the responsibility of graduate programs to provide these skills, not ignore them and turn out unqualified counselors.

The Assessment Course

What then should the traditional testing course now contain?

1. In the first place it should have as a prerequisite a basic course in statistical and measurement concepts. This basic course does not have to be at a particularly high level and can be taken at either the undergraduate or graduate levels as long as it provides an understanding of the normal curve, standard deviations, and correlational relationships. All counselors should have this knowledge not only for testing purposes, but even for understanding much of what appears in professional journals and in presentations at workshops and conventions. Without this prerequisite, a substantial portion of an appraisal course must be consumed with introducing simple measurement concepts, leaving insufficient time to cover the necessary

testing and appraisal skills. (Not to mention that negative attitudes are likely to be engendered in many counseling students if the first few weeks of a course in psychological testing are consumed with presenting various statistical concepts.)

2. A counseling appraisal course should contain the basic concepts of assessment including reliability and validity; standardized and non-standardized assessment techniques and performance measures; behavioral observation; and knowledge of and experience with some of the more common cognitive, career, and personality assessment instruments. The SAT/ACT and GRE/Miller can be cited as examples in illustrating some of these concepts, because almost all students have taken them and, as counselors, they will be expected to be knowledgeable about these ubiquitous instruments.

3. The testing course for counselors should differ from other testing courses in that it should emphasize the interpretation of test results. The importance of client interpretation and client understanding of the assessment results should be emphasized. The following principles should be stressed (Duckworth, 1990):

   a. Testing is carried out to generate information primarily for the benefit of the client and only to a lesser extent for the counselor's benefit.
   b. Clients should be active participants and collaborators in the testing process. As such, the client is involved in both the selection and the interpretation of tests.
   c. Clients can be assumed to be able to profit from the testing process if given appropriate feedback from the test results. They should receive an interpretation of the test results if this is possible. The interpretation should not use psychological jargon, and descriptions of behavior and feelings should be non-pejorative.
   d. Testing can give the total picture of an individual, and individuals need to know their weaknesses as well as their strengths. Assuming that the client is more normal than abnormal, the emphasis in test interpretation should be on normalcy, rather than on pathology.
   e. Most clients desire to change for the better and will do so if they can understand how change is possible. Testing is a useful tool to help clients see these possibilities.
   f. Work plays a very important role in people's lives, and vocational testing is often an important component of assessment.
   g. The ultimate goal of psychological testing in counseling is the empowerment of the client so that he or she can be more fulfilled through increased knowledge and skills.

4. Particular emphasis should be given to the use and interpretation of tests in counseling, with emphasis on individuals and individual differences.

5. Laboratory experience with the administration, scoring, and interpretation of various assessment instruments should also be included and introduced early. Students should take, score, and profile examples of different types of appraisal instruments. When possible, they should obtain some actual, practical experience in the administration and interpretation of assessment instruments (for example, by assisting undergraduates in a 101-level course).

6. In an ideal three-semester-hour assessment course, the counselor trainees would meet as a group for two hours a week and in separate sections for one hour, according to the type of counseling program in which they are involved. Marriage and family counselors would study marital and relationship inventories in their section; rehabilitation counselors topics such as the DSM-IV, work samples, vocational assessment, and evaluation of rehabilitation potential; mental-health counselors topics such as the assessment of depression, substance abuse, and the use of general mental-health inventories; and school counselors the assessment of achievement, academic aptitude, and the use of multi-aptitude batteries. Within each specialty section, relevant new, promising appraisal tools and methods of program assessment would be introduced. Such a course design is very difficult to implement but would solve many of the problems faced in teaching appraisal courses given the diversity of the counseling field. This modified course would concentrate on assessment concepts, leaving simulations and exercises to other courses where they are relevant—for example, genogram and family sculpturing exercises to the marital therapy course, and vocational card sorts and lifestyle exercises to the vocational/careers course.

In summary, the content of a course in assessment in counseling should consider the many types of tests and test results administered for other purposes but used by counselors, as well as those tests specifically employed in the counseling process. In the counseling setting, psychological tests are used to help clients to understand themselves. With the prevalence of negative attitudes toward psychological tests, counselors may be reluctant to make adequate use of them in assisting clients, but they should remember that the use of tests in counseling differs from test use in other settings. Counselors use tests primarily to assist individuals in developing their potential to the fullest and to their own satisfaction. Such results are not designed to be used by others to make decisions on clients' behalf; instead, they are to be used by clients themselves and only in those ways in which the clients decide to make use of them (Hood & Johnson, 1991). The emphasis must be on client understanding and client involvement in the assessment process.

References

Austin, J. T., & Hanisch, K. A. (1990). Occupational attainment as a function of abilities and interests: A longitudinal analysis using Project TALENT data. *Journal of Applied Psychology, 75*, 77–86.

Bradley, R. W. (1994). Tests and counseling: How did we ever become partners? *Measurement and Evaluation in Counseling and Development, 26*, 224–226.

Bubenzor, D. L., Zimpfer, D. G., & Mahrle, C. L. (1990). Standardized individual appraisal in agency and private practice: A survey. *Journal of Mental Health Counseling, 12*, 51–66.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., Hathaway, S. R., & McKinlay, J. C. (1989). *Minnesota Multiphasic Personality Inventory–2*. Minneapolis, MN: University of Minnesota Press.

Duckworth, J. (1990). The counseling approach to the use of testing. *The Counseling Psychologist, 18*, 198–204.

Elmore, P. B., Ekstrom, R. B., Diamond, E. E., & Whittaker, S. (1993). School counselors' test use patterns and practices. *The School Counselor, 41*, 73–80.

Elmore, P. B., Ekstrom, R. B., & Schafer, W. D. (1998, January). *Survey of assessment and evaluation activities of school counselors*. Paper presented at the Assessment '98 Conference of the Association for Assessment in Counseling, St. Petersburg, FL.

Engen, H. B., Lamb, R. R., & Prediger, D. J. (1982). Are secondary schools still using standardized tests? *Personnel and Guidance Journal, 60*, 287–290.

Goldman, L. (1972). Tests and counseling: The marriage that failed. *Measurement and Evaluation in Guidance, 4*, 197–205.

Goldman, L. (1992). Qualitative assessment: An approach for counselors. *Journal of Counseling and Development, 70*, 616–621.

Goldman, L. (1993, March). *Replacing the old testing course*. Paper presented at American Counseling Association convention, Atlanta, GA.

Goldman, L. (1994a). Marriage between tests and counseling redux: Summary of the 1972 article. *Measurement and Evaluation in Counseling and Development, 26*, 214–216.

Goldman, L. (1994b). The marriage is over . . . for most of us. *Measurement and Evaluation in Counseling and Development, 26*, 217–218.

Hills, J. R. (1991). Apathy concerning grading and testing. *Phi Delta Kappan, 72*, 540–545.

Hood, A. B., & Johnson, R. W. (1991). *Assessment in counseling: A guide to the use of psychological assessment procedures*. Alexandria, VA: American Association for Counseling and Development.

Impara, J. C., & Plake, B. S. (1995). Comparing counselors', school administrators', and teachers' knowledge in student assessment. *Measurement and Evaluation in Counseling and Guidance, 28*, 78–87.

Lowman, R. L., & Williams, R. E. (1987). Validity of self-ratings of abilities and competencies. *Journal of Vocational Behavior, 31*, 1–13.

Mabe, P. A. III, & West, S. G. (1982). Validity of self-evaluation of ability: A review of meta-analysis. *Journal of Applied Psychology, 67*, 280–296.

Prediger, D. J. (1994). Tests and counseling: The marriage that prevailed. *Measurement and Evaluation in Counseling and Development. 26*, 227–234.

Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan, 72*, 534–539.

Swanson, J. L. (1993). Integrated assessment of vocational interests and self-rated skills and abilities. *Journal of Career Assessment, 1*, 50–65.

Swanson, J. L., & Lease, S. H. (1990). Gender differences in self-ratings of abilities and skills. *Career Development Quarterly, 38*, 347–349.

Tennyson, W. W., Miller, G. D., Skovolt, T. B., & Williams, R. C. (1989). Secondary school counselors: What do they do? What is important? *The School Counselor, 36*, 253–259.

Watkins, C. E. Jr. (1991). What have surveys taught us about the teaching and practice of psychological assessment? *Journal of Personality Assessment, 56*, 426–437.

Watkins, C. E. Jr., Campbell, V. L., & Nieberding, R. (1994). The practice of vocational assessment by counseling psychologists. *The Counseling Psychologist, 22*, 115–128.

Zytowski, D. G., & Warman, R. E. (1982). The changing use of tests in counseling. *Measurement and Evaluation in Guidance, 15*, 147–152.

About the Author

Albert B. Hood is professor emeritus at the University of Iowa. He obtained his Ed.D. at Cornell University. Among his research interests are psychosocial development of college students and psychological assessment. Hood is the author of more than 100 publications including five books, among which are *Key Resources on Student Services* (Jossey-Bass) and *Assessment in Counseling* (American Counseling Association). He also serves as editor of the *Journal of College Student Personnel*. A Fellow of the APA, Hood is recipient of the Contribution to Knowledge Award (ACPA) and the Extended Research Award (ACA). He also was awarded a Fulbright Lectureship in the Soviet Union.

Chapter Six

The Pedagogical Basis for Multifaceted Assessment in Counselor Education

Barbara D. Yunker & Mary E. Stinson

Abstract

The teaching-learning process for counselors-in-training can be maximized through the use of multifaceted assessment. Extensive research shows the efficacy of involving students in making decisions about their own learning and assessment. The Jacksonville State University counselor-education program is used to exemplify how assessment results can become invaluable tools in the teaching-learning process.

The efficacy of using multifaceted assessment in education, rather than relying on a single measure for evaluating command of content domain and performance competencies, is virtually unchallenged today. Research substantiating this approach has dealt largely with issues related to validity and test bias. This paper examines another perspective, that of maximizing feedback from various forms of assessment in order to enhance the teaching-learning process in the training of new counselors. Such a review is timely, given the concerns raised in recent issues of *Counselor Education and Supervision* concerning the pedagogical foundation of counseling ("Restructuring," 2000; Sexton, 1998).

Student involvement and ownership are crucial in all aspects of effective education. This learning principle has been recognized (but not always practiced) since the beginning of the twentieth century when the Progressive Movement challenged the traditional, passive approach to education in which teachers lectured and students memorized content and recited by rote. John Dewey, who was perhaps the most articulate spokesperson for the movement, emphasized that students must be actively involved in their own education and that learning would be greatly enhanced by social interactions and a variety of experiences (Dewey, 1938). Scores of eminent theorists and researchers (Piaget, Vygotsky, Ausubel, Bloom, Bruner, and others) have substantiated the efficacy of active involvement on the part of the student. Their work on cognition initiated the Constructivist Movement in education, the theory that significant learning occurs only when the student finds the subject matter meaningful and the teaching-learning process is interactive and experiential.

Jerome Bruner, arguably the most influential contemporary learning theorist, made significant contributions in this regard. He proposed a discovery model of instruction that gave students the responsibility for choosing not only what they would learn, but how they would learn it. He added to our understanding of how meaningful curricula can be planned and assessment employed to enhance the learning process (Bruner, 1960, 1964, 1966, 1971). He described a spiral curriculum in which students would be introduced to concepts at an elementary level and later reintroduced to the concepts in various representations at progressively more complex and advanced levels. Right from the beginning, students would practice inquiry, self-monitoring, and self-correction, eventually evolving into self-motivated, autonomous learners. Bruner also demonstrated how assessment itself could become part of the instructional process. He described errors as hypotheses; i.e., responses that could be tested. Using Bruner's approach in the classroom gave students the opportunity to be active participants in both the learning and the evaluative phases of their education.

Another contemporary theorist, Howard Gardner, has added considerable impetus to the concept of student involvement in the teaching-learning process and to the important relationship between instruction and assessment (Gardner, 1983, 1991). Proponents of Gardner's theory of multiple intelligences recognize eight avenues by which students may solve a problem. Multiple ways of knowing require multiple ways of assessing that knowing. When assessment is directly linked to learning, it becomes a vital part of that learning. This link necessitates that students will assume some control over the evaluation procedures used to assess their skills, knowledge, and competencies. When students participate to this extent in the teaching-learning process, issues are probed more deeply, additional possibilities are explored, and significant responses emerge (Weber, 1999).

Students who are in the clinical sequence of their training need to feel a sense of empowerment in order to move toward professional autonomy (Nelson, 1997). Although becoming empowered may be a challenge for some students, particularly women, it is crucial that during their field experiences students assume a degree of authority within the supervisory relationship. Actively participating in the teaching-learning process through exercising a considerable degree of control over the assessment outcomes used to evaluate their knowledge and their performance paves the way for this empowerment to emerge and develop in students.

A cornerstone of training in counselor education is the full involvement of practicum and internship students in practicing the entire range of counselor roles and responsibilities and getting constructive feedback (Boylan, Malley, & Scott, 1995). Self-monitoring and self-correction are also critical. Students pull from their classroom experiences, as well as their life experiences, in order to participate fully in the supervised clinical experience. If those experiences have included collaboration, negotiation, and goal setting in the context of fulfilling a variety of assessment requirements, students are well on their way toward professional autonomy. All trainees, whatever their level, have reported being more satisfied with the supervisory relationship when they have had an active role in formulating their own goals (Nelson, 1997).

Instruction in the Jacksonville State University counselor-education program is geared to specified learning outcomes. The assessment instruments and procedures used are likewise matched to the types of learning outcomes being evaluated. Whatever the mode of instruction or type of assessment, student involvement and interaction are elicited. Applicable guidelines in the American Counseling Association *Code of Ethics and Standards of Practice* (1995) and the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999) are followed to insure that best practices are modeled. The general types of assessments used and the areas assessed are as follows:

1. Tests (objective and essay) are used to assess command of the knowledge base. To accompany forced-choice, paper-and-pencil test formats such as multiple-choice tests, we recommend an auxiliary assessment strategy termed "the Challenge" (Yunker, 1999). This posttest strategy provides a structured format for applying Bruner's theory that students' errors be treated as hypotheses to be tested (see the appendix).
2. Papers (research studies, literature reviews, critiques, etc.) are assigned to assess students' abilities to analyze, synthesize and organize information, and collect data, and to refine both research skills and writing mechanics.
3. Performance tasks accompanied by ratings and critiques (formal and informal classroom presentations, group projects, clinical skill demonstrations such as role-plays, audio and video tapes,

etc.) are used to assess communication and develop clinical skills.

4. Mini-portfolios (compiled during the counseling practicum and internship) assess and document a range of attributes and competencies expected of the emerging professional counselor. The mini-portfolios required in our counselor-education program include both core and student-selected products. Examples include audio and video tapes, group session plans, summaries or scripts of individual sessions, group participant evaluations, self-critiques, university and site supervisor critiques, and evaluations completed by all constituents.

We emphatically recommend student involvement in all phases of counselor training, including assessment. In order to make a determination concerning the current level of individualization in teaching and degree of student input, trainers could ask themselves the following questions:

- Do I employ multiple assessment techniques in each course I teach?
- Do I employ multiple assessment techniques in the clinical experiences I supervise?
- Is there compatibility between my instructional objectives and respective assessments?
- How do I use assessment results to provide feedback, instruction, and remediation?
- Is there more I could do to involve students in their own learning?
- How comfortable am I with students assuming more responsibility for their assessments and their subsequent empowerment as counselors?

It has become standard practice to use multifaceted assessment in counselor-education programs. In fact, the efficacy of this approach is supported in all types of educational evaluation that address validity and fairness issues. We hope we have highlighted another advantage of the multifaceted approach, one which may be underemphasized in the training of new counselors, namely how a variety of assessment results can become invaluable tools that enhance all aspects of the teaching-learning process.

In counselor-education programs, as in other educational programs, instruction is delivered and students are evaluated. There are several purposes for using a variety of assessments to evaluate students. According to Gronlund (1985), the purposes of assessment include (a) designing instructional objectives that reflect desired learning outcomes; (b) determining learners' needs; (c) providing relevant instruction based on assessment feedback; (d) evaluating intended outcomes; and (e) employing evaluation results to plan and improve educational programming. Our assessment techniques are formative and summative, assessing both the processes and products of learning. Formative evaluations—such as in-class activities, critical discussions, and role-plays—are employed to monitor student progress and provide continuous feedback (but not grades). Summative evaluations—in the form of tests, papers, presentations, and portfolios—assess the refined products of the teaching-learning process. Practicums and internships provide elements of both. Using such a multifaceted approach to assessment permits counselors-in-training to evaluate their learning of theoretical concepts, and it enables them to demonstrate their critical thinking, their powers of persuasion, and their creativity. Active involvement in the assessment processes fosters students' confidence in their ability to become effective practicing counselors.

References

American Counseling Association. (1995). *Code of ethics and standards of practice*. Alexandria, VA: Author.

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.

Boylan, J. C., Malley, P. B., & Scott, J. (1995). *Practicum and internship: Textbook for counseling and psychotherapy* (2d ed.). Washington, DC: Accelerated Development.

Bruner, J. (1960). *The process of education*. Cambridge, MA: Harvard University.

Bruner, J. (1964). The course of cognitive growth. *American Psychologist, 19*, 1–15.

Bruner, J. (1966). *Toward a theory of instruction*. Cambridge, MA: Harvard University Press.

Bruner, J. (1971). *The relevance of education*. New York: Norton.

Dewey, J. (1938). *Experience and education*. New York: Macmillan.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Gardner, H. (1991). *The unschooled mind: How children think and how schools should teach.* New York: Basic Books.

Gronlund, N. E. (1985). *Measurement and evaluation in teaching* (5th ed.). New York: Macmillan.

Nelson, M. L. (1997). An interactional model for empowering women in supervision. *Counselor Education and Supervision, 37*, 125–139.

Restructuring clinical training: In pursuit of evidence-based practice. (2000). [Special section]. *Counselor Education and Supervision, 39*, (4) .

Sexton, T. L. (Section coordinator). (1998). Reconstructing counselor education: Issues of our pedagogical foundation [Special section]. *Counselor Education and Supervision, 38*, (2).

Weber, E. (1999). *Student assessment that works: A practical approach.* Boston: Allyn and Bacon.

Yunker, B. (1999). Adding authenticity to traditional multiple-choice test formats. *Education, 120*, 82–87.

About the Authors

Barbara D. Yunker is an associate professor in the Department of Educational Resources and program chair of secondary education at Jacksonville State University, Jacksonville, Alabama. She obtained her Ph.D. in educational psychology with a specialty in school psychology from the University of Alabama and has worked as a teacher, behavior specialist, psychometrist, and school psychologist. Yunker has published articles and book reviews in the *Journal of Reading, Personnel Psychology, Behavior Disorders, Education,* and ERIC documents. Her current research interests include gender equity and assessment issues.

Mary E. Stinson is an associate professor emerita in the College of Education, Jacksonville State University, Jacksonville, Alabama, where she continues to teach a limited number of courses and supervise internships in the Counselor Education Program. She holds an Ed.S. in counselor education and a Ph.D. in educational psychology from the University of Alabama. Her primary areas of expertise are crisis counseling, counseling supervision, and assessment. She has made scores of presentations on these topics at area, state, and national conferences and has extensive experience as a consultant and supervising LPC.

Appendix

The Challenge is a highly structured discussion between students and a resident expert (the instructor) following the return of graded test papers that encourages students to question the answer key to some multiple-choice test items. When test items are constructed beyond the knowledge level of Bloom's cognitive domain (i.e., comprehension, application, analysis, and synthesis), there may be room for a logical defense in support of another answer choice. It is our contention that students who are able to provide an objective and rational justification for an answer choice that differs from the keyed target should get credit for that answer. This strategy turns the written portion of the test into an instructional tool and reinforces retention of useful information. A synopsis of the Challenge method follows:

1. Choose or develop multiple-choice test items to reflect the course objectives at all levels in the cognitive domain of Bloom's taxonomy.
2. When introducing the test, instruct students to choose what they perceive to be the best answer to each item.
3. Administer and score the test according to the answer key.
4. Return graded papers or test booklets to students and set aside about an hour (depending on the length of the test) for the Challenge activity.
5. Read each test item with its keyed answer aloud to the class. Instruct them to identify those items they might want to question when you are finished.
6. Spell out the rules for the activity and enforce them strictly.
   - Students must raise a hand and wait to be recognized before initiating a challenge. This rule teaches patience and self-control.
   - Students must phrase the challenge with "I" statements not "you" statements. For example, students could say "I interpreted number __ to mean __, so I chose __ as the best answer"; or, "For number __ , I chose __ because __."
   - Students must defend and justify their answer choices rationally. This rule encourages articulate communication and accountability.
   - Students must be objective. No whining or hostility is tolerated. Students learn to disagree in a constructive,

nonviolent way.

- Peers are encouraged to provide support for the challenge or for the key. No put-downs are allowed. This practice encourages cooperation and reinforces retention of the pertinent information.
- As the instructor, you reserve the right to accept or reject a challenge without argument from the students.

7. Indicate how students should mark successfully challenged test items to receive credit.
8. Require that students return all test papers and answer sheets to you. Students who keep a test paper automatically receive a grade of F.
8. If you accept a challenge, all students who chose the same tag *and* were present for the Challenge activity also receive credit. Absentees are not eligible for credit for successfully challenged test items. This creates motivation to participate.

K–12 Assessment and Counseling

Addressing Fragmentation: Building Integrated Services for Student Support
Jackie M. Allen

Competencies in Assessment and Evaluation for School Counselors
Joint Committee of the American School Counselor Association (ASCA) and the Association for Assessment in Counseling (AAC), chaired by Patricia Elmore

Assessing the Effectiveness of School Guidance Programs: Program, Personnel, and Results Evaluation
Norman C. Gysbers

Chapter Seven

Addressing Fragmentation: Building Integrated Services for Student Support

Jackie M. Allen

Abstract

The concept of integrated service delivery has been endorsed by the professional organizations of all student support services personnel. Yet this paradigm has been slow to be adopted in schools. Barriers to implementation are discussed, along with potential solutions to those barriers. Finally, the process by which a school or district might move toward integrated service delivery is outlined.

With higher benchmarks for student achievement, more required assessment, and increased demand for accountability, pupil personnel programs and services are often cut when budgets are trimmed. Yet, more money may be only part of the answer to improving programs and services. Fragmentation in services for children and youth and the lack of collaboration to develop effective service models may be as large a barrier as the lack of funds. The print component of a two-hour teleconference, Investing in Our Youth: A Nationwide Committee of the Whole, was devoted to addressing the fragmentation in services for children and youth. In this document the need for coordination and collaboration was described: "The current system of fragmented services for youth has reached the limit of its effectiveness, and even at its peak, such a system fails to meet the complex needs of today's youth" (Palaich, Whitney, & Paolino, 1991, p. v).

Integrated services—that is, programs based on a collaborative model provided by credentialed pupil personnel professionals—are not a new concept. In 1994 the theme of the American School Counselor Association (ASCA) annual conference was "School Counselors Collaborating for Student Success." The first ASCA Presidential Theme Digest developed from this conference outlined the impending educational issues; characteristics, requirements, and benefits of collaboration; and the collaborative role of the school counselor in educational reform. In 1996, a statement in the Mental Health in Schools Center newsletter summarized the need for collaboration and the lack of response by professionals to this need: "In both policy and practice, it is evident that developing a comprehensive, integrated approach is a low priority" (p. 2). Implicit in California's comprehensive school health model is the necessity for collaborative efforts in order to have a coordinated school health program in a local school (California Department of Education, 2000).

The National Alliance of Pupil Service Organizations (NAPSO), a coalition of national professional organizations whose members provide a variety of student support services and programs, endorses a collaborative approach to the delivery of services and programs to meet the complex needs of the nation's youth. In 1994 the California Alliance of Pupil Service Organizations adopted a position statement, School-Linked, School-Based Integrated Services, which embraced the collaborative model of school-based, school-linked integrated service delivery and asserted its essential role in meeting the increasingly complex needs of California's children and their families. In *School Psychology: A Blueprint for Training and Practice II* the National Association of School Psychologists advocates for the use of a collaborative/participatory model in mental health service and program delivery (Ysseldyke et al., 1997). School nurses and school social workers have also supported a shared agenda and integrated service delivery (Gibelman, 1993; National Association of School Nurses, n.d.).

Collaboration is widely recognized by national professional associations and pupil service organizations, is recommended by national studies on the delivery of youth services, and is a basic concept of the comprehensive school health model. Why is it such an extremely difficult concept to actualize at the local school level? Are we cheating our students by not providing collaborative, coordinated services? What role does the student support professional play in the coordination of student support programs and services?

Overcoming Barriers

Perhaps the place to start in addressing this problem is an overview of the barriers to team building that appear to exist in our schools. An initial concern is territoriality and turf issues. Each specialist or professional may perceive that another professional is taking over his or her role or unique responsibilities when, in fact, there is more than enough work to go around for everyone. If the primary concern of a student support program is the student, then the most important goal is to serve the student, and perhaps who provides the service is not as important as the accomplishment of the goal. An effective student support program will increase, not decrease, the need for personnel.

Fragmented, categorical funding may lead to divisiveness among student support staff. For example, special education has a designated source of resources through state and federal funding and may be viewed as having program and job security. Such categorical funding is not adequate to meet the needs of all students, and limited resources become the much larger concern. Creative use of funding sources and the creation of new funding through special grants will improve student support programs. Collaborative legislative efforts, responsible assessment and accountability, and social marketing campaigns will increase visibility and financial support.

Student support professionals (counselors, psychologists, social workers, and nurses) are not always aware of the distinct roles of each member of an integrated services staff. Staff training, beginning in graduate education programs, is essential for specialists to obtain a viable perspective of the whole picture of support services. Through the Integrating Pupil Services Personnel Into Comprehensive School Health and HIV Prevention grant from the Centers for Disease Control and Prevention, and with the administrative efforts of the Education Development Center, Inc., five national pupil services personnel organizations—the American School Counselor Association, National Association of School Nurses, National Association of School Psychologists, American Psychological Association, and National Association of Social Workers—collaborated to strengthen the roles of the professions they represent in comprehensive school health activities at local, state, and national levels. One of the major initial tasks of the grant was to develop a training model of integrated services to demonstrate how the various professions would work together on a school-site student study team to address the needs of students. In order to prepare the presentation, it was necessary to determine the shared roles and unique contributions each professional brought to the team. The effort of struggling with the common and unique roles of various disciplines provided each specialist with an understanding of both the whole picture and each professional's special contribution. Each specialist approaches student needs from a slightly different perspective—the nurse from a health perspective, the social worker from a family systems and ecological perspective, the psychologist from a learning theory and assessment approach, and the school counselor from an academic, social/personal, and career emphasis. Yet student support teams work together and share common roles in educational reform, program planning, crisis intervention and prevention, community support building, and assessment and referral from the perspective of whole-child development and with the ultimate goal of school and community wellness.

Disjointed organizational structure may be a significant barrier to team building. In the school-district-level designation of coordination, supervision, and accountability, nurses are separated from counselors, counselors from psychologists, and social workers from other student support staff members. Therefore, it is very difficult to develop clear lines of communication and a coherent policy for pupil personnel services and programs. The support staff needs to communicate with each other not just at IEP meetings, but at times when they can plan a coordinated, comprehensive program to address student needs in the district and at the local school site. The concept of a comprehensive school health program is a model for uniting eight diverse components of school health under one umbrella. Most schools do not have the resources or personnel to implement all eight components in one comprehensive program. Uniting student support staff to work collaboratively in coordinated efforts to improve pupil service programs is a step we must take. Such an effort will provide the support students need to be healthy in mind and body, achieve academically, develop satisfying relationships, and prepare for responsible citizenship and the world of work.

Fear of change may impede team building in a district or at a local school site. Collaboration implies change: forming new service delivery models, looking at service delivery in new ways, seeking and adopting new paradigms, and challenging both oneself and the system. In a popular management book *Who Moved My Cheese?* Johnson (1998) reminds us through his parable that we all react to change in different ways but that those who "hem and haw," refusing to accept the challenges of change, may never find the cheese and may not be able to work effectively in the system. Breaking down the barrier of fear of the unknown is crucial for the change process.

Collaborative efforts will lay the groundwork for developing coherent policies and clear goals. An important collaborative effort in every school is the disaster plan, which specifies what, who, when, where, and how all personnel and students in the school should function in the event of a disaster. Since Columbine and other school tragedies, more attention has been given to a wide variety of possible crisis situations requiring the awareness, knowledge, and combined efforts of all staff to maintain student safety. Student support professionals need to make the development of prevention and intervention plans and programs a top priority for their collaborative efforts.

Developing Integrated Service Programs

Many benefits may be derived from collaboration. Student support personnel, credentialed or certified school counselors, school psychologists, school social workers, and school nurses provide support services and programs for students in our nation's schools. Together student support professionals are able to create a united front in legislative, public relations, and program development. Understanding the issues of stakeholders, which may be an overwhelming task for one profession, becomes much easier with collaborative efforts. Coordinated efforts create increased visibility, reduce turf competition, and increase the amount and scope of services. Collaborative programs are more cost effective because integrated services staff share experiences, concerns, and ideas and thus increase their individual awareness and knowledge of what their colleagues do in their specialized jobs.

Envisioning the future is the beginning of change. At a local school site, student support professionals need to meet together, focus on the needs of their students, and develop a shared vision. The process of developing a shared vision is the first important step. This vision might be based on a comprehensive school health model (California Department of Education, 2000), the ASCA standards (Dahir, Sheldon, & Valiga, 1998), a comprehensive counseling and guidance model (Gysbers and Henderson, 2000), or a unique integrated services model created locally. Agonizing over turf issues, program design, diminishing resources, duplication or gaps in services, and the overwhelming demands of meeting student needs often builds dynamic relationships between student support personnel.

The creation of a collaborative work culture where professionals spend time together doing strategic program planning enhances the change process. The planning process must include all parties affected by student support services, including students, parents, administrators, teachers, all student support personnel, and representatives from the community. Employing diverse modalities such as singing, recreational activities, art, and drama in the planning process improves the development of a collaborative work culture and the possibility of designing an effective program plan. An impartial facilitator in the planning process may help to keep the lines of communication open, to assist in the definition of roles, and to promote creativity in decision making.

Beginning with a needs assessment of the school climate and community will assure that members of a planning team know the strengths and weaknesses of the existing services, what is important to school and community members, and the specific needs of students. Scanning for economic, political, and other external environmental indicators can help determine the major emphases to be included in the program. Using surveys and questionnaires, existing evaluations, and both informal and formal feedback will clarify the challenges to be faced in the collaborative effort to make school better for kids.

Now the real work to solidify the vision into reality begins with translation of the needs into a plan of action. Short-term and long-term goals are determined based on the needs assessment. Strategies are developed to carry out the goals. Resources are analyzed and, when necessary, additional resources are sought. The roles of the student support personnel must be clarified, and an evaluation component should be built into the model.

Finally, support for the new integrated services model is sought and the stakeholders in the process begin a public relations campaign to announce the changes and gain support for the new model. It is advisable to institutionalize the changes made in the program or services model in order to guarantee permanent progress. The school community needs to be aware of the programmatic changes and the benefits to be gained by those changes. A successful public relations campaign will lead to a successful change in the program and services paradigm.

Student support personnel can be significant catalysts for collaboration and change at their schools by facilitating a culture of collaboration in student services and programs and by developing integrated services models that meet the needs of students and the school community. Fragmentation in children's and youth services will disappear when the stakeholders and service providers meet to discuss their community's needs and concerns. Barriers to team building can be surmounted and integrated service models developed. It is imperative that educators form partnerships with parents, staff, and community so they can bring together the necessary resources to support students in realizing academic self-esteem, academic achievement, and school-to-work readiness. A paradigm of change is possible. Collaboration is the key to moving student support programs into the twenty-first century and providing the quality of services our nation's youth deserve.

References and Resources

Allen, J. (1994). *School counselors collaborating for student success.* (ASCA Presidential Theme Digest). Greensboro, NC: ERIC Clearinghouse on Counseling and Student Services. (ERIC Document Reproduction Service No. EDO-CG-94-27)

California Alliance of Pupil Service Organizations. (1994). School-linked, school-based integrated services. (Position Statement). Unpublished.

California Department of Education (2000). *Building infrastructure for coordinated school health: California's blueprint*. Sacramento, CA: Author.

Dahir, C., Sheldon, C, and Valiga, M. (1998). *Vision into action: Implementing the national standards for school counseling programs*. Alexandria, VA: American School Counselor Association.

Gibelman, M. (1993). School social workers, counselors, and psychologists in collaboration: A shared agenda. *Social Work in Education 15*(1), 45–51.

Gysbers, N., & Henderson, P. (2000). *Developing and managing your school guidance program*. Alexandria, VA: American Counseling Association.

Johnson, S. (1998). *Who moved my cheese?* New York: G. P. Putnam's Sons.

Marx, E., & Wooley, S. (1998). *Health is academic: A guide to coordinated school health programs*. New York: Teachers College Press.

National Association of School Nurses (n.d.). *Integrated service delivery*. (NASN Issue Brief). Scarborough, ME: Author.

Palaich, R., Whitney, T., & Paolino, A. (1991). *Changing delivery systems: Addressing the fragmentation in children and youth services* [Booklet]. Columbia: South Carolina ETV.

Policies and practices for addressing barriers to student learning: Current status and new directions. (1996). *Addressing Barriers to Learning, 1*(4), 1–2.

Smith, R. C., Lincoln, C., & Dodson, D. (1991). *Let's do it our way: Working together for educational excellence* [Booklet]. Columbia: South Carolina ETV.

Ysseldyke, J., et al. (1997). *School psychology: A blueprint for training and practice II*. Bethesda, MD: National Association of School Psychologists.

About the Author

Jackie M. Allen is an education programs consultant with the Student Support Services and Programs unit of the California

Department of Education. She received her doctorate in education at the University of San Francisco in 1990. Her primary research interests are counseling and the special needs student, assessment and accountability of student support services, and counseling and the human brain. The editor of *School Counseling: New Perspectives and Practices,* Allen has also authored numerous articles on counseling and student support services. She is a past president of the American School Counselor Association, has served two terms on the American Counseling Association Governing Council, and has received the California Association of Counseling and Development Clarion Modell Distinguished Service Award.

Chapter Eight

Competencies in Assessment and Evaluation for School Counselors

Approved by the American School Counselor Association
on September 21, 1998, and by the Association for
Assessment in Counseling on September 10, 1998[1]

Patricia Elmore, Chair

Abstract

The purpose of these competencies is to provide a description of the knowledge and skills that school counselors need in the areas of assessment and evaluation. Because effectiveness in assessment and evaluation is critical to effective counseling, these competencies are important for school counselor education and practice. Although consistent with existing Council for Accreditation of Counseling and Related Educational Programs (CACREP) and National Association of State Directors of Teacher Education and Certification (NASDTEC) standards for preparing counselors, they focus on competencies of individual counselors rather than content of counselor-education programs.

The competencies can be used by counselor and assessment educators as a guide in the development and evaluation of school counselor preparation programs, workshops, in-services, and other continuing-education opportunities. They may also be used by school counselors to evaluate their own professional development and needs for continuing education.

School counselors should meet each of the nine numbered competencies and have the specific skills listed under each competency.

Competency 1. *School counselors are skilled in choosing assessment strategies.*

a. They can describe the nature and use of different types of formal and informal assessments, including questionnaires, checklists, interviews, inventories, tests, observations, surveys, and performance assessments, and work with individuals skilled in clinical assessment.

b. They can specify the types of information most readily obtained from different assessment approaches.

c. They are familiar with resources for critically evaluating each type of assessment and can use them in choosing appropriate assessment strategies.

d. They are able to advise and assist others (e.g., a school district) in choosing appropriate assessment strategies.

Competency 2. *School counselors can identify, access, and evaluate the most commonly used assessment instruments.*

a. They know which assessment instruments are most commonly used in school settings to assess intelligence, aptitude, achievement, personality, work values, and interests, including computer-assisted versions and other alternate formats.

b. They know the dimensions along which assessment instruments should be evaluated, including purpose, validity, utility, norms, reliability and measurement error, score reporting method, and consequences of use.

c. They can obtain and evaluate information about the quality of those assessment instruments.

Competency 3. *School counselors are skilled in the techniques of administration and methods of scoring assessment instruments.*

a. They can implement appropriate administration procedures, including administration using computers.

b. They can standardize administration of assessments when interpretation is in relation to external norms.

c. They can modify administration of assessments to accommodate individual differences consistent with publisher recommendations and current statements of professional practice.

d. They can provide consultation, information, and training to others who assist with administration and scoring.

e. They know when it is necessary to obtain informed consent from parents or guardians before administering an assessment.

Competency 4. *School counselors are skilled in interpreting and reporting assessment results.*

a. They can explain scores that are commonly reported, such as percentile ranks, standard scores, and grade equivalents. They can interpret a confidence interval for an individual score based on a standard error of measurement.

b. They can evaluate the appropriateness of a norm group when interpreting the scores of an individual or a group.

c. They are skilled in communicating assessment information to others, including teachers, administrators, students, parents, and the community. They are aware of the rights students and parents have to know assessment results and decisions made as a consequence of any assessment.

d. They can evaluate their own strengths and limitations in the use of assessment instruments and in assessing students with disabilities or linguistic or cultural differences. They know how to identify professionals with appropriate training and experience for consultation.

e. They know the legal and ethical principles about confidentiality and disclosure of assessment information and recognize the need to abide by district policy on retention and use of assessment information.

Competency 5. *School counselors are skilled in using assessment results in decision making.*

a. They recognize the limitations of using a single score in making an educational decision and know how to obtain multiple sources of information to improve such decisions.

b. They can evaluate their own expertise for making decisions based on assessment results. They also can evaluate the limitations of conclusions provided by others, including the reliability and validity of computer-assisted assessment interpretations.

c. They can evaluate whether the available evidence is adequate to support the intended use of an assessment result for decision making, particularly when that use has not been recommended by the developer of the assessment instrument.

d. They can evaluate the rationale underlying the use of qualifying scores for placement in educational programs or courses of study.

e. They can evaluate the consequences of assessment-related decisions and avoid actions that would have unintended negative consequences.

Competency 6. *School counselors are skilled in producing, interpreting, and presenting statistical information about assessment results.*

a. They can describe data (e.g., test scores, grades, demographic information) by forming frequency distributions, preparing tables, drawing graphs, and calculating descriptive indices of central tendency, variability, and relationship.

b. They can compare a score from an assessment instrument with an existing distribution, describe the placement of a score within a normal distribution, and draw appropriate inferences.

c. They can interpret statistics used to describe characteristics of assessment instruments, including difficulty and discrimination indices, reliability and validity coefficients, and standard errors of measurement.

d. They can identify and interpret inferential statistics when comparing groups, making predictions, and drawing conclusions needed for educational planning and decisions.

e. They can use computers for data management, statistical analysis, and production of tables and graphs for reporting and interpreting results.

Competency 7. *School counselors are skilled in conducting and interpreting evaluations of school counseling programs and counseling-related interventions.*

a. They understand and appreciate the role that evaluation plays in the program development process throughout the life of a program.

b. They can describe the purposes of an evaluation and the types of decisions to be based on evaluation information.

c. They can evaluate the degree to which information can justify conclusions and decisions about a program.

d. They can evaluate the extent to which student outcome measures match program goals.

e. They can identify and evaluate possibilities for unintended outcomes and possible impacts of one program on other programs.

f. They can recognize potential conflicts of interest and other factors that may bias the results of evaluations.

Competency 8. *School counselors are skilled in adapting and using questionnaires, surveys, and other assessments to meet local needs.*

a. They can write specifications and questions for local assessments.

b. They can assemble an assessment into a usable format and provide directions for its use.

c. They can design and implement scoring processes and procedures for information feedback.

Competency 9. *School counselors know how to engage in professionally responsible assessment and evaluation practices.*

a. They understand how to act in accordance with ACA's *Code of Ethics and Standards of Practice* and ASCA's *Ethical Standards for School Counselors.*

b. They can use professional codes and standards, including the *Code of Fair Testing Practices in Education, Code of Professional Responsibilities in Educational Measurement, Responsibilities of Users of Standardized Tests,* and *Standards for Educational and Psychological Testing,* to evaluate counseling practices using assessments.

c. They understand test fairness and can avoid the selection of biased assessment instruments and biased uses of assessment instruments. They can evaluate the potential for unfairness when tests are used incorrectly and for possible bias in the interpretation of assessment results.

d. They understand the legal and ethical principles and practices regarding test security, copying copyrighted materials, and unsupervised use of assessment instruments that are not intended for self-administration.

e. They can obtain and maintain available credentialing that demonstrates their skills in assessment and evaluation.

f. They know how to identify and participate in educational and training opportunities to maintain competence and acquire new skills in assessment and evaluation.

Definitions of Terms

Competencies describe skills or understandings that a school counselor should possess to perform assessment and evaluation activities effectively.

Assessment is the gathering of information for decision making about individuals, groups, programs, or processes. Assessment targets include abilities, achievements, personality variables, aptitudes, attitudes, preferences, interests, values, demographics, and other characteristics. Assessment procedures include but are not limited to standardized and unstandardized tests, questionnaires, inventories, checklists, observations, portfolios, performance assessments, rating scales, surveys, interviews, and other clinical measures.

Evaluation is the collection and interpretation of information to make judgments about individuals, programs, or processes that lead to decisions and future actions.

About the Committee Chair

Patricia B. Elmore is associate dean for administrative services in the College of Education and professor in the Department of Educational Psychology and Special Education at Southern Illinois University, Carbondale. She received her Ph.D. in 1970 in educational psychology with a specialization in educational measurement and statistics. She is a past president of the Association for Assessment in Counseling and the 1994 recipient with Ruth B. Ekstrom and Esther E. Diamond of the American Counseling Association Research Award. Currently she is editor of *Measurement and Evaluation in Counseling and Development*. She co-authored *Basic Statistics* with Paula L. Woehlke (Addison Wesley Longman, 1997) and has published in *Educational and Psychological Measurement, Educational Researcher, Journal for Research in Mathematics Education, Journal of Educational Measurement, Journal of Educational Psychology, Measurement and Evaluation in Counseling and Development*, and *The School Counselor*, among other journals.

1. A joint committee of the American School Counselor Association (ASCA) and the Association for Assessment in Counseling (AAC) was appointed by the respective presidents in 1993 with the charge to draft a statement about school counselor preparation in assessment and evaluation. Committee members were Ruth Ekstrom (AAC), Patricia Elmore (AAC, Chair, 1997–1999), Daren Hutchinson (ASCA), Marjorie Mastie (AAC), Kathy O'Rourke (ASCA), William Schafer (AAC, Chair, 1993–1997), Thomas Trotter (ASCA), and Barbara Webster (ASCA).

Chapter Nine

Assessing the Effectiveness of School Guidance Programs: Program, Personnel, and Results Evaluation

Norman C. Gysbers

Abstract

In order to fully evaluate comprehensive school guidance programs, three forms of evaluation are required. First, the program must be reviewed using program standards, evidence, and documentation to establish that a written guidance program exists in a school district or building and that the written program matches the implemented program. Second, guidance-program personnel need job descriptions derived directly from the program so that evaluation forms can be developed and used for formative and summative personnel evaluation. Third, results evaluation that focuses on the impact of the guidance and counseling activities in the guidance curriculum, individual planning, responsive services, and system-support components of a comprehensive guidance program is necessary. The results of 20 years of research show positive effects of effective guidance counseling on students' academic achievement.

Demonstrating accountability through the measured effectiveness of the delivery of guidance programs and the performance of the guidance personnel involved helps ensure that students, parents, teachers, administrators, and the general public will continue to benefit from quality, comprehensive guidance programs. To achieve accountability, evaluation is needed concerning the nature, structure, organization, and implementation of school-district guidance programs; the school counselors and other personnel who are implementing the programs; and the impact the programs are having on students, the schools where they learn, and the communities in which they live. This means that the overall evaluation of school-district guidance programs should be approached in the following three ways: program evaluation, personnel evaluation, and results evaluation (Gysbers & Henderson, 2000). This article defines each of these types of evaluation and then briefly describes how each type of evaluation can be carried out. Finally, the last section presents data from a number of studies highlighting what we have learned so far from results evaluation efforts.

Program Evaluation

Program evaluation measures two questions: Does the school district have a written comprehensive guidance program? Is the written program of the district being implemented fully in the school buildings of that district? Answers to these questions are provided through a process called *program evaluation*, the goal of which is to examine the written program carefully and verify through documentation that it is the program being implemented. Whether or not a written guidance program exists in the district and whether or not any discrepancies exist between the written guidance program and the program actually implemented become clear as the program evaluation process unfolds.

To conduct program evaluation, program standards are required. *Program standards* are acknowledged measures of comparison or the criteria used to make judgments about the adequacy of the nature and structure of the program as well as the degree to which the program is in place. How many program standards are required to establish whether a comprehensive guidance program is in place and functioning? The answer is that sufficient standards are required to ensure that judgments can be made concerning whether or not a complete, comprehensive guidance program is actually in place and functioning to a high enough degree to benefit fully all students, parents, teachers, and the community. To illustrate what a program standard looks like, here is an example:

> The school district is able to demonstrate that all students are provided the opportunity to gain knowledge, skills, values, and attitudes that lead to a self-sufficient, socially responsible life.

A school district meeting this standard has defined the content that all students should learn in a systematic, sequential way. The content goals are tied to those defined in the basic mission of the school district and are based on human development theories regarding individuals' personal, social, career, and educational development. The content is further defined in a scope and sequence that outlines the guidance curriculum. The implementation of the guidance-curriculum component of a comprehensive guidance program entails teaching lessons and units designed to help students acquire the competencies outlined in the scope and sequence.

What would an evaluator look for to see that this standard is in place? Here are some examples:

- A developmentally appropriate guidance curriculum that teaches all students the knowledge and skills they need to be self-sufficient and lead socially responsible lives.
- A guidance curriculum that is articulated from elementary to middle to high schools.
- Priorities that are established for the acquisition of competencies by students at each grade level or grade-level grouping.
- Sufficient curriculum materials to support the teaching of the needed knowledge and skills.
- A yearly schedule that incorporates classroom guidance units.
- Students in special education and other special programs receiving guidance curriculum instruction.

To make judgments about guidance programs using standards, evidence is needed concerning whether or not the standards are being met. In program evaluation such evidence is called *documentation.* For the standard listed previously, evidence that it is in place might include the following:

- District guidance-curriculum guides.
- District guidance-curriculum scope and sequence.
- Teachers' and counselors' lesson plans.
- Yearly master calendar for the guidance program.

(Gysbers & Henderson, 2000, p. 405)

Personnel Evaluation

A key part of comprehensive guidance program implementation and management is a school counselor performance improvement system.[1] The basic purpose of this system is to assist school counselors in reaching and enhancing their professional potential. It helps individuals define their jobs, provide professional supervision, conduct performance evaluation, and set goals for continued professional development. The purposes of evaluating school counselors' performance are to improve the delivery to and impact of the program on the students and parents it serves and to provide for communication among school counselors, guidance-program staff leaders, and school administrators. For school counselors, evaluation specifies contract status recommendations and provides summative evaluation as to their effectiveness. For the school district, evaluation defines expectations for school counselors' performance and provides a systematic means of measuring their performance relative to these expectations.

The three facets of the performance-evaluation part of a school counselor performance-improvement system are (a) self evaluation, (b) administrative evaluation, and (c) assessment of goal attainment. Self-evaluation and administrative evaluation focus on job-performance competencies and represent data-supported professional judgments as to school counselors' proficiency in using the skills and commitment levels required for their jobs. The assessment of goal attainment focuses on school counselors' efforts to improve the program and their professionalism.

For performance evaluation to be done fairly, many data sources are used as each part of a performance-improvement system is implemented. Specific examples of typical behaviors of individual school counselors are gathered throughout the year and documented. These patterns of behavior are then compared and contrasted with clearly stated professional standards. Recently the state of Missouri adopted a set of standards for professional school counselor evaluation. These standards with criteria are as follows (Missouri Department of Elementary and Secondary Education, 2000, pp. 27–28):

Standard 1: The professional school counselor implements the Guidance Curriculum Component through the use of effective instructional skills and the careful planning of structured group sessions for all students.

Criterion 1: The professional school counselor teaches guidance units effectively.

Criterion 2: The professional school counselor encourages staff involvement to ensure the effective implementation of the guidance curriculum.

Standard 2: The professional school counselor implements the Individual Planning Component by guiding individuals and groups of students and their parents through the development of educational and career plans.

Criterion 3: The professional school counselor, in collaboration with parents, helps students establish goals and develop and use planning skills.

Criterion 4: The professional school counselor demonstrates accurate andappropriate interpretation of assessment data and the presentation of relevant, unbiased information.

Standard 3: The professional school counselor implements the Responsive Services Component through the effective use of individual and small-group counseling, consultation, and referral skills.

Criterion 5: The professional school counselor counsels individual students and small groups of students with identified needs/concerns.

Criterion 6: The professional school counselor consults effectively with parents, teachers, administrators, and other relevant individuals.

Criterion 7: The professional school counselor implements an effective referral process in collaboration with parents, administrators, teachers, and other school personnel.

Standard 4: The professional school counselor implements the System Support component through effective guidance program management and support for other educational programs.

Criterion 8: The professional school counselor provides a comprehensive and balanced guidance program in collaboration with school staff.

Criterion 9: The professional school counselor provides support for other school programs.

Standard 5: The professional school counselor uses professional communication and interaction with the school community.

Criterion 10: The professional school counselor demonstrates positive interpersonal relations with students.

Criterion 11: The professional school counselor demonstrates positive interpersonal relations with educational staff.

Criterion 12: The professional school counselor demonstrates positive interpersonal relations with parents/patrons.

Standard 6: The professional school counselor fulfills professional responsibilities.

Criterion 13: The professional school counselor demonstrates a commitment to ongoing professional growth.

Criterion 14: The professional school counselor possesses professional and responsible work habits.

Criterion 15: The professional school counselor follows the profession's ethical and legal standards and guidelines, as well as promotes cultural diversity and inclusivity in school policy and interpersonal relationships.

Results Evaluation

Having established that a guidance program is operating in a school district through program evaluation, and having established through personnel evaluation that school counselors and other guidance program personnel are carrying out the duties listed on their job descriptions 100% of the time, it now is possible to evaluate the results of the program. Johnson (1991) suggested that there are long-range, intermediate, immediate, and unplanned-for results that need consideration. According to Johnson, *long-range results* focus on how programs affect students after they have left school. Usually long-range results are gathered using follow-up studies. *Intermediate results* focus on the knowledge and skills all students may gain by graduation from participating in the guidance program. *Immediate results* are the knowledge and skills students may gain from participating in specific guidance activities. Finally, the possibility of *unplanned-for results* that may occur as a consequence of guidance activities also needs to be taken into account.

For the purposes of this article, illustrations of immediate and intermediate results evaluation using the structure of the Missouri Comprehensive Guidance Program (Gysbers, Starr, & Magnuson, 1998) are presented in the form of two research questions. First, do students master guidance competencies as a result of their participation in the guidance curriculum component of the program (immediate evaluation)? Second, do students develop and use career plans as a result of their participation in the individual planning component of the program (intermediate evaluation)?

Immediate Evaluation: Guidance Competency Mastery

Do students master guidance competencies? Johnson (1991) outlined the following procedures to answer this question in terms of immediate results. First the competencies to be mastered need to be identified. Second *what* results (what students should be able to write, talk about, or do) are specified. Then *who* will conduct the evaluation is decided. This is followed by a design about *when* the evaluation is done. Then *criteria* are established so that judgments can be made about students' mastery of guidance competencies. Finally, *how* all of this is to be accomplished is specified.

Another way to conduct immediate evaluation to measure mastery of guidance competencies is the use of a confidence survey. In this format, guidance competencies are listed and students are asked to rate on a Likert scale how confident they are that they have mastered these competencies. The confidence survey can then be used as a pre-post measure. Gain scores can be obtained and related to such measures as academic achievement and vocational identity (Gysbers, Lapan, Multon, & Lukin, 1992; Lapan, Gysbers, Hughey, & Arni, 1993).

Intermediate Evaluation: Career Plans

Do students develop and use career plans? In making judgments concerning the career plans of students, criteria need to be identified as to what makes a good plan. Four criteria are recommended: a plan needs to be comprehensive, developmental, student-centered and student-directed, and competency-based. One way to evaluate students' career plans is to judge the extent to which the activities included in the individual planning component of the guidance program lead to the development of plans that meet these criteria. A second way is to make judgments about the adequacy of the plans' contents. Finally, a third way is to judge their use. Do students actually use their career plans in planning for the future?

What Have We Learned So Far From Results Evaluation?

The major reason to plan, design, and implement comprehensive guidance programs is to assist students in their academic, career, and personal development, working in close consultation with their parents. Do guidance programs and the interventions used produce measurable results? The cumulative empirical research evidence from more than 20 years of professional literature unequivocally indicates that the answer to this question is yes.

What kind of results do guidance programs and the interventions used produce? Here are some examples. In a major review of the literature in school counseling, Borders and Drury (1992) found that guidance-program interventions have a substantial impact on students' educational and personal development and that they contribute to students' success in the classroom. Gerler (1985) analyzed a decade of research on the results of elementary school counseling and found that guidance-program interventions in the affective, behavioral, and interpersonal domains of students' lives positively affected students' academic achievement. The results of a study by Lee (1993) showed that classroom guidance lessons in elementary school led by school counselors positively influenced students' academic achievement in mathematics. Similar results were found by St. Clair (1989) in her review of the impact of guidance-program interventions at the middle school level. Further, Evans and Burck (1992) conducted a meta-analysis of 67 studies concerning the impact of career education interventions (career guidance) on students' academic achievement. Their results supported the value of these interventions as contributors to the academic achievement of students.

More recently, studies conducted in Missouri and Utah provide additional evidence of the value of comprehensive guidance programs. In a study conducted in Missouri high schools, Lapan, Gysbers, and Sun (1997) found that students in high schools with more fully implemented guidance programs were more likely to report that they had earned higher grades, their education was better preparing them for their future, their school made more career and college information available to them, and their school had a more positive climate. In another study in Missouri, when classroom teachers in 184 small-, medium-, and large-sized middle schools rated guidance programs in their schools as more fully implemented, seventh graders in these schools reported that they had earned higher grades, school was more relevant for them, they had positive relationships with teachers, they were satisfied with their education, and they felt safer in school (Lapan, Gysbers, & Petroski, in press). In addition, in a study conducted in the state of Utah, strong guidance programs were found effective in helping students target areas of educational or career emphasis. In schools with highly implemented programs there were also documented increases in enrollment for courses related to specific educational goals or careers, e.g., advanced math and science courses and vocational/technical courses. In addition, high student performance on the American College Test (ACT), a standardized achievement test, was related to enrollment in schools with highly implemented guidance programs. Scores were significantly higher on all four skill areas of the ACT (mathematics, English, reading, and science) than student scores from low-implementing schools and the scores for the state of Utah as a whole. These results suggest student learning increases when courses are organized around a relevant area of interest (Utah State Office of Education, 2000).

Finally, in their review of outcome research in school counseling, Sexton, Whiston, Bleuer, & Walz (1997, p. 125) made the following points:

- Reviews of outcome research in school counseling are generally positive about the effects of school counseling.
- Research results do indicate that individual planning interventions can have a positive impact on the development of students' career plans.
- There is some support for responsive service activities such as social skills training, family-support programs, and peer counseling. Consultation activities are also found to be an effective school counseling activity.

References

Borders, L. D., & Drury, S. M. (1992). Comprehensive school counseling programs: A review for policy makers and practitioners. *Journal of Counseling and Development, 70*, 487–498.

Evans, J. H., & Burck, H. D. (1992). The effects of career education interventions on academic achievement: A meta-analysis. *Journal of Counseling and Development, 71*, 63–68.

Gerler, E. R. (1985). Elementary school counseling research and the classroom learning environment. *Elementary School Guidance and Counseling, 20*, 39–48.

Gysbers, N. C., & Henderson, P. (2000). *Developing and managing your school guidance program* (3rd ed.). Alexandria, VA: American Counseling Association.

Gysbers, N. C., Lapan, R. T., Multon, K. D., & Lukin, L. (1992). *Missouri guidance competency evaluation survey, grades 4–6, 6–9, and 9–12*. Columbia, MO: Center for Educational Assessment.

Gysbers, N.C., Starr, M., & Magnuson, C. (1998). *Missouri comprehensive guidance: A model for program development, implementation, and evaluation* (Rev. ed.). Jefferson City, MO: Missouri Department of Elementary and Secondary Education.

Henderson, P., & Gysbers, N. C. (1998). *Leading and managing your school guidance program staff*. Alexandria VA: American Counseling Association.

Johnson, C. D. (1991). Assessing results. In S. K. Johnson & E. A. Whitfield (Eds.), *Evaluating guidance programs*. Iowa City, IA: American College Testing Program.

Lapan, R. T., Gysbers, N. C., Hughey, K., & Arni, T. J. (1993). Evaluating a guidance and language arts unit for high school juniors. *Journal of Counseling and Development, 71*, 444–451.

Lapan, R. T., Gysbers, N. C., & Petroski, G. F. (In press). Helping seventh graders be safe and successful: A statewide study of the impact of comprehensive guidance programs. *Journal of Counseling and Development*.

Lapan, R. T., Gysbers, N. C., & Sun, Y. (1997). The impact of more fully implemented guidance programs on the school experiences of high school students: A statewide evaluation study. *Journal of Counseling and Development, 75*, 292–302.

Lee, R. S. (1993). Effects of classroom guidance on student achievement. *Elementary School Guidance and Counseling, 27*, 163–171.

Missouri Department of Elementary and Secondary Education (2000). *Guidelines for performance-based professional school counselor evaluation*. Jefferson City, MO: Author.

Sexton, T. L., Whiston, S. C., Bleuer, J. C., & Walz, G. R. (1997). *Integrating outcome research into counseling practice and training*. Alexandria, VA: American Counseling Association.

St. Clair, K. L.(1989). Middle school counseling research: A resource for school counselors. *Elementary School Guidance and Counseling, 23*, 219–226.

Utah State Office of Education. (2000). *1989–1999: A decade of progress and change in the delivery of comprehensive guidance programs, grades 7–12*. Salt Lake City, UT: Author.

About the Author

Norman C. Gysbers is a professor in the Department of Educational and Counseling Psychology at the University of Missouri–Columbia. He received his M.A. (1959) and Ph.D. (1963) degrees from the University of Michigan. Gysbers' research and teaching interests are in career development; career counseling; and school guidance and counseling program development, management, and evaluation. He is author of 59 articles in 17 different professional journals, 25 chapters in published books, 14 monographs, and 13 books, including *Career Counseling: Process, Issues, and Techniques*, 1998 (with Mary Heppner and Joseph Johnston); *Leading and Managing Your School Guidance Program Staff*, 1998 (with Patricia Henderson); and *Developing and Managing Your School Guidance Program* (3rd. ed.), 2000 (with Patricia Henderson). In 1981 he was awarded the National Vocational Guidance Association National Merit Award and in 1983 the American Counseling Association Distinguished Professional Service Award. In 1989 he received the National Career Development Association Eminent Career Award and in 2000 he received the National Career Development Association Presidents Recognition Award. Currently Gysbers is the editor of the *Journal of Career Development*.

1. For a complete description and discussion of a school counselor performance-improvement system, please see Henderson, P., & Gysbers, N. C. (1998). *Leading and managing your school guidance program staff*. Alexandria VA: American Counseling Association.

Special Populations

Assessing Diverse Populations
Courtland C. Lee

Equity Issues in the Assessment of Individuals With Visual or Hearing Impairments
Ruth B. Ekstrom

Modifying Tests for Students With Disabilities
Douglas K. Smith

Chapter Ten

Assessing Diverse Populations

Courtland C. Lee[1]

Abstract

At the beginning of the twenty-first century multiculturalism and diversity are major challenges and opportunities. Assessment must be considered in a cultural context, the basis of which is worldview. Failure to take such dynamics into consideration could affect not only the assessment process but also the interpretation of assessment information. A list of changes that are necessary in the assessment process is provided.

I must confess that preparing my article was both a very easy and yet an extremely challenging task. It was easy in that I can address issues of multiculturalism and diversity in my sleep. Yet it was challenging, because I kept asking myself: "What can I say about assessment and diversity that has not already been said in all of those chapters at the end of our testing books, the ones that are usually titled, "Assessing Special Populations?"

Still, in the words of the old Bob Dylan classic, "The Times, They Are a Changin'!" I would like to start by having you engage in a little thought about the concept of changing times. Each evening my wife and I watch the news while eating our dinner. Usually, as we watch incredible news stories from around the country night after night, we find ourselves shaking our heads and saying, "That's life in America at the end of the twentieth century." Consider the following headlines and news flashes that reflect life in America at the end of the twentieth century:

- The OJ verdict splits the country into two seemingly different countries: one Black, one White.
- The Million Man March and the Promise Keepers Rally bring empowered men to the Mall in Washington.
- Ellen "comes out" on national television.

- The name of a school in Louisiana is changed because it was named after a famous slave owner: George Washington
- Texas and California end their affirmative action programs.
- Three out four African American males in Washington, D.C., are either in jail, waiting to be sentenced, or on probation.
- The Americans With Disabilities Act changes the face of American architecture.
- A young boy in Chicago is beaten into a coma because he was the wrong color in the wrong neighborhood at the wrong time.
- A university president finds himself in hot water for his "Oreo" remark.
- President Clinton calls for a Dialogue on Race, but it gets bogged down in politics and idle rhetoric.

Each of these headlines and news flashes underscores how profoundly the issues of multiculturalism and diversity impact our consciousness at the end of the twentieth century. At the beginning of the twentieth century, W. E. B. DuBois wrote that the "problem of America is the problem of the color line." I think that at the beginning of the twenty-first century, we can paraphrase DuBois' quote and say that the challenge of America is multiculturalism and diversity. The color line is still very much present, but it has been extended to include other areas of cultural difference, such as gender, sexual orientation, physical disability, and social class.

However, as we move forward in the twenty-first century, I think that multiculturalism and diversity are the great promise of America as well. Let me share with you how the field of assessment can play a part in realizing that promise. Here are some thoughts and ideas about changes in assessment that can lead to assessment for change in a culturally diverse society.

Assessment as a process must be considered within a *cultural context*. The basis for understanding cultural context is the concept of *worldview*. A worldview is how, over time and over the world, people have come to view the many facets of the human experience. A very important point to consider is that there are major differences in how people come to view the world. Consider, for example, the fundamental differences in how people view language. Language is integral to assessment. Language *is* culture. Languages are not different words for the same thing; languages are different words for entirely different ways of seeing and conceptualizing the world. The words we use largely determine how we perceive the world. Because of differences in aspects of worldview such as language, people present their personalities, cognitive abilities, interests and other psychoeducational constructs differently—not better or worse just differently!

Cultural differences in how people come to view the human experience beg the fundamental question. "Can we develop assessment tools that accurately and fairly assess psychoeducational constructs across cultures?" Consider the cultural contexts and worldviews of two very different people:

Ronald

Ronald is a nine-year-old African American male in the fourth grade in an elementary school in southeast Washington, D.C. He comes to school every day proudly wearing the latest fashions, including a cap, baggy/saggy pants, and expensive sneakers with the laces untied. Most days when he comes into the classroom, his female teacher confronts him about removing his cap, pulling up his pants, and tying his shoelaces. Ronald usually storms over to the other side of the room, mumbles under his breath, and grudgingly removes his cap.

As the morning's instruction proceeds, Ronald occupies his time interacting with the other boys who sit around him. He enjoys talking with them, giving them "high fives," and generally joking with and teasing them. The teacher perceives Ronald to be inattentive and the instigator of most of this activity. She proceeds to reprimand him about his behavior.

When the teacher reprimands Ronald, he gets upset at her protestations, claiming that she is picking on him. She claims that he is not paying attention. However, when she presses him about the topic under class discussion, he is able to respond correctly. In fact, Ronald claims that he has raised his hand several times that morning, but that the teacher has ignored him.

The teacher notices again that Ronald's shoelaces are still untied. She sternly orders him to tie the laces. Ronald staunchly refuses, stating that this is the way they are supposed to be worn. She states that in her classroom, shoelaces will be tied. Again, she orders him to lace the shoes and moves toward Ronald, placing her hand on his shoulder and looking him squarely in the face.

At this point, Ronald jerks away from the teacher and shouts, "Don't be touchin' me!" He forcefully walks away from the teacher, picks up a book and flings it across the classroom. The teacher then orders Ronald to go to the principal's office.

Sarah

Sarah is a 30-year-old White woman from a close-knit rural mountain community in central Virginia. She is married with two small children. Sarah has reluctantly left her husband because he has severely beaten her. When she left home, she took refuge at a women's center in a nearby town.

Meeting with a counselor at the women's center, Sarah expresses strong fundamentalist Christian beliefs. When the counselor asks about her facial bruises, she states that her husband often beats her to ensure that she remains a "good Christian woman." When asked to consider the possibility of divorcing her husband, Sarah states that she cannot do so as this would be considered a sin. She claims she would lose her children and be shunned by her family and friends if she took such an action. Sarah expresses concern about her lack of employment experience. She says that the Bible mandates that her place is at home caring for her husband and children.

Let's suppose Ronald is mandated for psychoeducational assessment and Sarah's counselor suggests that she go for career assessment. What are some of concepts related to cultural context that we would want to consider in these cases? What are some of the dynamics that influence the worldview of Ronald and of Sarah that could affect both the assessment process and outcome? The dynamics of language, kinship, religion/spirituality, roles and status, sex role socialization, learning style, and attitudinal orientation are readily apparent. Likewise, environmental factors such as racism, sexism, and economic disadvantage appear to have influenced the psychosocial development of these two individuals.

Failure to take such dynamics into consideration could affect not only the assessment process but also the interpretation of assessment information as well. Because important decisions are predicated on the outcome of the assessment process, there are several changes that must be considered in assessment if assessment is to be used for change:

1. Ensuring that the development of multicultural/diversity competencies are an integral part of the development of assessment competencies. We need culturally responsive assessment professionals.
2. Clearing up the dichotomy between culture-specific and culture-fair assessment techniques.
3. Ensuring that normative samples reflect multiculturalism and diversity to the fullest extent possible.
4. Ensuring fair access to assessment—not only for test

administrators but for test takers as well.

5. Ensuring that no decision about an individual is made on the basis of a single test score. All relevant information about an individual, including his or her worldview and cultural context, must be part of the decision-making process.
6. Redefining concepts of validity in light of increasing diversity.
7. Adopting a twenty-first century agenda with respect to assessment and multiculturalism.

In closing, I think that collectively as a profession, we counselors can lead the way in confronting the challenges and promoting the promise of cultural diversity in the twenty-first century. We can be catalysts for positive social change. That idea was my message as ACA president in 1997–98. Specifically, those of you who specialize in assessment have the power to promote the important process for collecting decision-making data in important new, culturally responsive ways. I would urge you to use your powers for good. For example, work to erase the stigma and suspicion about your specialty that lingers in many diverse communities.

As you continue to construct assessment instruments and adopt testing standards, let the wisdom of an older African American woman from the 1950s guide your thinking. This wisdom comes from Lena Younger, the matriarch in Lorraine Hansberry's classic American play "Raisin in the Sun" which remains the quintessential view of African American family life:

> "When you starts to measure somebody, measure 'em right child, measure 'em right. Make sure you done taken into account what hills and valleys he done come through before he gets to wherever he is."

That should be the context of assessment for change. That's life in America at the start of the twenty-first century.

About the Author

Courtland C. Lee is dean of the School of Education and professor of counselor education at Hunter College of the City University of New York. He received his Ph.D. in counseling from Michigan State University. Lee is the author, editor, or co-editor of four books on multicultural counseling and the author of two books on counseling African American males. In addition, he has published numerous book chapters and articles on counseling across cultures. Lee is a past president of the American Counseling Association, of the Association for Multicultural Counseling and Development, and of

Chi Sigma Iota, the international counseling honor society.

1. This article was the keynote address at the Assessment '98: Assessment for Change—Changes in Assessment conference, St. Petersburg, FL, January 16–18, 1998.

Chapter Eleven

Equity Issues in the Assessment of Individuals With Visual or Hearing Impairments

Ruth B. Ekstrom[1]

Abstract

Legislation such as the Americans With Disabilities Act mandates testing accommodations, both in the manner of presentation of the examination and in access to the testing site, in order to ensure the examination accurately reflects the abilities of a person with a disability. Individualized testing accommodations and test aids are described, including changes to the test directions, administration procedures, test content, or means of response, audio-taped examinations, or interpreters. Issues regarding whether modified tests are equivalent to standard tests are presented, focusing specifically on the low reading levels of most people with hearing impairment. Finally, legal issues, such as the inclusion of students with disabilities in national testing programs and voluntary disclosure of a disability are explained.

Providing equitable assessment of individuals with visual or hearing impairments, whether for rehabilitation, education, employment, clinical, or counseling purposes, presents a number of challenges. First of all, it is critical that the assessment reflect the abilities of the individual, not the disability. This is may be done through the use of tests specifically designed for people with disabilities. Alternatively, test accommodations or modifications may be made to standardized tests. It is important to note, however, that not all individuals with disabilities require special tests or testing accommodations.

Tests designed specifically for assessing individuals who are blind or visually impaired include cognitive instruments, such as the Blind Learning Aptitude Test, and developmental rating scales, such as the Maxfield-Buchholz Social Maturity Scale for Blind Pre-School

Children. But only a small proportion of the tests most commonly used with blind and partially sighted individuals were developed for this population (Swallow, 1981). Even when tests have standardization specific to visually impaired individuals, this may be based on unreplicated norms from small, possibly biased samples (such as students in residential institutions for the blind) or may be based on a non-homogeneous sample of individuals with various types and degrees of visual impairment (Simeonsson, 1986). Similar instruments, with similar problems, exist for individuals with hearing impairment.

Providing testing accommodations for individuals with disabilities is nothing new; in 1937 the College Board developed a version of the SAT for visually impaired students. Today, because of legislation such as the Americans With Disabilities Act (ADA; PL 101-336), much more attention is being given to test accommodations. For example, ADA says:

> Any private entity that offers examinations or courses related to applications, licensing, certification, or credentialing for secondary or postsecondary education, professional, or trade purposes . . . must assure that when the examination is selected and administered to an individual with a disability . . . the examination results accurately reflect the individual's aptitude or achievement level or whatever other factor the examination purports to measure, rather than reflecting the individual's impaired sensory, manual, or speaking skills (except where those skills are the factors that the examination purports to measure).

Testing accommodations for individuals with visual or hearing impairment may involve changes in the test directions and administration procedures, changes in the test content, and changes in test response mechanisms. ADA specifically states that modifications to an examination may include changes in the length of time for completion and adaptation of the manner in which the examination is given. Provision of appropriate auxiliary aids is also required under ADA, "unless offering a particular auxiliary aid would fundamentally alter the measurement of the skills or knowledge the examination is intended to test or would result in an undue burden."

Test aids and services mentioned in ADA include "taped examinations, interpreters or other means of making orally delivered materials available to individuals with hearing impairments, Braille or large-print examinations and answer sheets or qualified readers for individuals with visual impairments or learning disabilities, transcribers for individuals with manual impairments, and other similar services and actions." Other accommodations for individuals with visual impairments may include provision of special lighting; magnification devices; tactile maps, diagrams, and graphs; audiocassettes; electronic readers (speech synthesizers); and talking calculators. Individuals with hearing disabilities may use video cassettes, especially those using sign language translations of the test directions or test content. Siskind (1993a) has described the modifications used in statewide testing programs to accommodate pupils with disabilities.

ADA mandates that test accommodations be individualized. Equitable does not mean identical. No single type of test accommodation is adequate or appropriate for all individuals with a given type of visual or hearing impairment. Often more than one type of accommodation is required by an individual test taker. For example, an individual with visual impairment may need both an audio-taped version of the test and a large-type "follow along" script; another test taker may need a Braille version of a test and tactile maps and graphs along with extra time. (The typical Braille reader may require 2 to 2 1/2 times as long to read material as would a sighted individual, but this time frame may vary considerably both because of the nature of the test material and because of the individual's skill in reading Braille.)

Sometimes the content of a test must be changed. Such adaptations might include, for hearing impaired individuals, dropping the listening comprehension part of a foreign language test. The question then arises as to whether or not the test modification is appropriate. It is very important to consider the construct being measured and to determine whether the testing accommodation or modification alters that construct. For example, a student using a large-print version of a reading comprehension test is still reading, but a student using an audiocassette version of the test is displaying skills in listening comprehension, not reading comprehension.

The rationale for the testing accommodation or modification must be carefully considered. When testing individuals with visual impairment, the report from a functional visual assessment can provide important information in this regard. Such assessments describe how the individuals use their vision. (It is important to remember that more than 75% of individuals classified as legally blind have some usable vision.) This report may indicate the type of lighting needed to optimize use of vision, the most appropriate type size, the best posture for individuals with a limited field of vision, optimum distance for viewing material, and recommendations for using low-vision equipment. If you are doing assessments for rehabilitation or educational planning, or for clinical purposes, this information is critical, and you should request it prior to carrying out any testing. If you are testing individuals for admissions or employment, however, the law prohibits your making an inquiry about the existence of a disability prior to making the admissions decision or job offer. This is confusing to many people. I want to emphasize that you should know the purpose of testing before you make an inquiry about a disability.

If you are working with teachers or special educators on assessments that are part of mandatory state testing programs for all students, be sure that you and they know the testing accommodations and modifications that are allowed. One study (Siskind, 1993b) found that "neither special or regular educators are well informed about this topic." I have special concerns about the validity of test scores when a student with a disability is not able to use a test accommodation that she or he is familiar with and that has been requested.

Most of the research on test accommodations shows that the modified tests are comparable to the standardized versions. The studies in the book *Testing Handicapped People* (Willingham et al., 1988) compared test results based on such measures as reliability, validity, factor structure, and prediction of academic performance. In general, comparability between nonstandard and standard test administrations was high. But both this research and other research done at ACT (Laing and Farmer, 1984) suggest that the prediction of grades for students with physical disabilities is somewhat less accurate than for other students. It should also be pointed out, however, that the Educational Testing Service studies showed that visually impaired students performed slightly better than expected.

Modified tests may not always be equivalent to standard forms. For example, an audio-taped version of a test places much more emphasis on memory skills than does a print version. Certain mathematical item types tend to present more difficulties for students using the Braille version of tests, especially when the items contain graphical material or where spatial estimation can be helpful in eliminating options. Charts, graphs, and diagrams also may present special problems for test takers who are visually impaired.

For individuals with severe hearing impairments, use of any verbal test may be problematic due to their limited English language skills (Gordon, Stump, and Glaser, 1996). In this country, deaf individuals and those with severe hearing impairments, especially those whose hearing loss occurred before they acquired speech, often communicate using American Sign Language (ASL)—which has a different grammar and syntax than English—and learn English as a second language. For this reason, some individuals have argued that instruments such as the Test of English as a Second Language (TOEFL) might be more appropriate for assessing the verbal skills of deaf students than an instrument such as the SAT (Ragosta & Nelson, 1986; Traxler, 1990). The difficulties of assessing individuals with hearing impairments are not limited to tests of verbal ability, however. An interest inventory, a personality scale, or any other test that requires a sixth- to eighth-grade reading level may be invalid for many individuals in this population. The mean reading level for people with hearing impairment has been estimated at the third- to fourth-grade level (Schmelter-Davis, 1984). Individuals who experience hearing loss in their adult years may try to rely on lip reading, but even skilled lip readers understand only about 25% of what is being said (Vernon & Andrews, 1990). Because hearing impairments, unlike most other physical disabilities, are invisible to others, test administrators may have difficulty in determining whether the test taker understands what is being said.

Remember that some individuals with visual and hearing impairments may try to conceal them. (See, for example, the book *Planet of the Blind* by Stephen Kuusisto, in which the author describes how he tried for nearly four decades to hide the fact that he was legally blind.) Blind individuals often develop exceptional memory skills, particularly to help themselves with orientation and mobility. Test data shows that students who are blind tend to have better short-term memory skills than the general population. Individuals with hearing impairment may also try to conceal their disability. It has been estimated that it takes an average of seven years for someone with a hearing impairment to seek help and that one out of every seven individuals with a hearing impairment never seeks help.

In addition to making appropriate and individualized testing accommodations, equity requires that the testing site be accessible to individuals with disabilities. ADA requires that examinations be offered "in a place and manner accessible to persons with disabilities" or that alternative arrangements be made. Alternative arrangements mentioned in ADA include "provision of an examination in an individual's home with a proctor if accessible facilities or equipment are unavailable." One good source of information about administering tests to individuals with disabilities is *Guide for Administering Written Employment Examinations to Persons with Disabilities* (Eyde, Nestor, Heaton, & Nelson, 1994). It is important to provide a testing site that is free of obstacles and, for individuals with visual impairment, to orient the test taker to the test room. Test takers should be informed in advance about any aids that will be used and be told whether they may bring any aids with them. Orientation to the aids used in the testing situation may be necessary, even if the test taker uses similar aids at home, in school, or in the workplace.

Access to state and national testing programs has been a special concern for students with disabilities. The Individuals With Disabilities Education Act of 1991 requires that most students with disabilities be included in district, state, and national assessments. Despite this requirement, many of these students have been excluded from such programs. For example, in 1994, the National Assessment of Educational Progress (NAEP) included only 50% of grade 4 students, 38% of grade 8 students, and 36% of grade 12 students who were identified as having an Individualized Educational Plan. (These plans are required for students with disabilities who demonstrate a need for special education and services.) Under the new procedures, students with an IEP will be included unless the IEP team determines that the student cannot participate or the student's cognitive functioning is so severely impaired that she or he cannot participate, even with accommodations (Olson and Goldstein, 1996). As has been pointed out by staff at the National Center on Educational Outcomes, exclusion of students with disabilities from state and national testing programs limits our ability to obtain policy-relevant information on educational outcomes for this population and perpetuates the myth of inherent differences (McGrew, Thurlow, Shriner & Spiegel, 1992).

In competitive situations such as admissions and employment, equity demands that applicants not be asked to reveal possibly prejudicial information about the existence of a disability prior to receiving the admission or job offer. Section 504 of the Rehabilitation Act of 1973 prohibits test score recipients from making preadmission inquiries as to whether or not an applicant has a disability. Applicants can be "invited" to reveal a disability, but they must be told that the information is being requested on a voluntary basis and will be kept confidential. This legal requirement raises serious problems, especially with nationally standardized admission tests. Many test score recipients feel a need to know whether a test was given under nonstandard conditions and, thus, may be less valid. Currently, the U.S. Office of Education Office of Civil Rights (OCR) has an "interim" policy that postsecondary institutions may use test scores that indicate the test was taken under nonstandard conditions if the test score is not the only criterion for admission and the individual is not denied admission because she or he took the test under nonstandard conditions.

Finally, it is critical that test administrators and test interpreters do not hold biased or stereotyped views about individuals with disabilities. Testing professionals have the responsibility to become informed about disabilities and to correct any misconceptions they hold about the capabilities of individuals with disabilities. The emerging field of disability studies (see the *Chronicle of Higher Education*, January 23, 1998) is providing a body of literature with valuable insights into disability experiences.

References

Bradley-Johnson, S. (1994). *Psychoeducational assessment of students who are visually impaired or blind: Infancy through high school* (2nd ed.). Austin, TX: PRO-ED.

Eyde, L. D., Nester, M. A., Heaton, S. M., & Nelson, A. V. (1994). Guide for administering written employment examinations to persons with disabilities (PRDC-94-111). Washington, DC: U.S. Office of Personnel Management, Personnel Research and Development Center.

Gordon, R. P., Stump, K, & Glaser, B. A. (1996). Assessment of individuals with hearing impairments: Equity in testing procedures and accommodations. *Measurement and Evaluation in Counseling and Development, 29*(2), 111–118.

Harris, L. K., VanZant, C. E., & Rees, T. H. (1997). Counseling needs of students who are deaf and hard of hearing. *The School Counselor, 44*(4), 271–279.

Kuusisto, S. (1997). *Planet of the blind.* New York: Dial Press.

Laing, J., & Farmer, M. (1984). Use of the ACT assessment by examinees with disabilities (ACT Research Rep. No. 84). Iowa City, IA: American College Testing Publications.

McGrew, K. S., Thurlow, M. L., Shriner, J. G., & Spiegel, A. N. (1992). Inclusion of students with disabilities in national and state data collection programs (Technical Report 2). Minneapolis, MN: National Center on Educational Outcomes, University of Minnesota.

Olson, J. G., & Goldstein, A. A. (1996). Increasing the inclusion of students with disabilities and limited English proficiency students in NAEP. (NCES 96-894). *Focus on NAEP, 2*(1).

Ragosta, M., & Nelson, C. (1986). *TOEFL and hearing impaired students: A feasibility study* (Technical Report No 143). (ERIC Document Reproduction Service No. ED 275-422).

Schmelter-Davis, L. (1984). *Vocational evaluation of handicapped college students: Hearing, motor, and visually impaired.* (ERIC Document Reproduction Service No. ED 264 390)

Sherman, S., & Robinson, N. (Eds.) (1982). *Ability testing of handicapped people: Dilemma for government, science, and the public.* Washington, DC: National Academy Press.

Simeonsson, R. J. (1986). *Psychological and developmental assessment of special children.* Boston, MA: Allyn & Bacon.

Siskind, T. G. (1993a). Modifications in statewide criterion-referenced testing programs to accommodate pupils with disabilities. *Diagnostique, 18*(3), 232–249.

Siskind, T. G. (1993b). Teachers' knowledge about test modifications for students with disabilities. *Diagnostique, 18*(2), 145–157.

Swallow, R. (1981, February). Fifty assessment instruments commonly used with blind and partially sighted individuals. *Visual Impairment and Blindness,* 65–72.

Traxler, C. B. (1990). *Direct writing assessment of non-traditional students: Construct validity of the TOEFL-TWE.* Paper presented at the American Educational Research Association annual meeting, Boston, MA.

Vernon, M., & Andrews, J. F. (1990). *The psychology of deafness.* New York: Longman.

Willingham, W. W., Ragosta, M., Bennett, R. E., Braun, H., Rock, D. A., & Powers, D. E. (1988). *Testing handicapped people.* Needham Heights, MA: Allyn and Bacon.

Zieziula, F. R. (Ed.). (1986). *Assessment of hearing-impaired people: A guide for selecting psychological, educational, and vocational tests.* Washington, DC: Gallaudet University Press.

About the Author

Ruth Burt Ekstrom is a principal research scientist in the Higher Education Research Division of Educational Testing Service. Her work at ETS has included studies of test use, guidance and counseling, student achievement, tracking, and women's education and employment. Ekstrom holds a master's degree from Boston University and a doctorate from Rutgers University. She is co-author of the book *Education and American Youth: The Impact of the High School Experience* and of numerous book chapters. Ekstrom has served on the Joint Committee on Testing Practices and the American Psychological Association Committee on Psychological Tests and Assessment. In 1996 Ekstrom received the American Counseling Association Extended Research Award for her career accomplishments.

1. This article was originally presented in the Symposium: Test Interpretation and Diversity: Achieving Equity in Assessment, at the Assessment '98: Assessment for Change—Changes in Assessment conference, St. Petersburg, FL, January 16–18, 1998.

Chapter Twelve

Modifying Tests for Students With Disabilities

Douglas K. Smith[1]

Abstract

Modifying standardized tests for students with disabilities is a complex issue. Tests should be modified only when alternative measures do not exist. Testing professionals should always be cognizant of the fact that whenever modifications are made, normative interpretations should be made very cautiously. In addition, the accommodations that were made should be described and the examiner should continually ask whether the accommodations significantly alter the format of the test or change the nature of the test. In this paper, many issues related to test modification are highlighted and a step-by-step procedure for developing appropriate testing accommodations is presented.

Providing testing accommodations for individuals with disabilities is not a new concept. Accommodations have been required within the educational setting since the passage of Public Law 94-142 (The Education for All Handicapped Children Act) and within the public setting since passage of the Rehabilitation Act of 1973. Many accommodations, such as Braille, large print, and extra time, have become common. In considering testing accommodations, we usually think of accommodations needed to assist individuals with physical or sensory disabilities. However, recent legislation (the Americans With Disabilities Act and the 1997 Individuals With Disabilities Act Amendments) has expanded our definitions of both disabilities and testing accommodations. For the first time, students with disabilities are to be included in state and district testing programs (unless specifically excluded from such testing in their individualized educational plan) and necessary accommodations are to be provided. Of course, not all students with disabilities require accommodations, and accommodations may not be needed for all assessments. The need for such accommodations must be determined on a case-by-case basis by considering the student involved and the specific nature and purpose of assessment.

The number of reasons for assessing students with disabilities continues to expand. Assessment is mandated for placement in special education programs, and periodic re-evaluations are required for developing individualized educational plans. The emphasis on educational accountability has resulted in more and more district- and state-mandated assessments. Assessment is also utilized in planning transitional services for students with disabilities and in rehabilitation program planning.

Accommodations reflect changes in the standard or usual way in which a test is administered so that a student with a disability is not penalized by the disability. In other words, the accommodations are designed to "level the playing field" and to insure that we are measuring the student's abilities, not disabilities. Testing accommodations may involve changes in the setting, timing, scheduling, presentation, or response required on the test (Thurlow, Elliott, & Ysseldyke, 1998). Legislation requires that testing be conducted in settings that are physically accessible to the individual being tested. Examples of changes in the setting may include special lighting or testing in a separate room. The focus of this paper, however, is the process for making accommodations to individually administered, standardized tests. It is assumed that testing will occur in an appropriate environment accessible to the student.

Modifications to timing may include providing the student with additional time to complete the test, eliminating bonus points for rapid performance, allowing additional exposure time for test stimuli, providing frequent breaks, or allowing unlimited time. Scheduling modifications may include changing the order in which subtests are administered, testing over an extended period of time rather than in one sitting, or testing only at specific times of day. Changes in presentation mode may involve the use of sign language, large print, Braille, or repetition of directions. Response modifications may involve responding verbally instead of in writing, or using a word processor instead of writing, for example. In general, accommodations for physical or sensory disabilities are less problematic than accommodations for cognitive or affective disabilities because the latter may be less apparent to the examiner but of equal importance and impact on the individual (Olson & Goldstein, 1997).

The accommodations made for a disability may have a substantial impact on the subsequent scores obtained and may affect the validity of those scores. Some types of accommodations may be appropriate in some situations but not in others. How is one to decide whether an accommodation is appropriate? What factors should be considered in developing appropriate accommodations? Does the purpose of the testing affect the appropriateness of specific accommodations? These are some of the questions that are examined in this paper. In this paper I provide a procedure or process for making testing accommodations. Although each situation in which a testing accommodation may be needed is unique and should be treated individually, there are some universal principles or guidelines that form the basis for the decisions that we make.

As testing professionals, we are guided by the ethical standards of our professional organizations as well as relevant state and federal laws. Perhaps none is more influential than the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999). In the latest edition of the standards, an entire chapter is devoted to the assessment of individuals with disabilities. The chapter addresses some of the more common types of accommodations, situations in which accommodations may and may not be appropriate, and possible effects of accommodations on test scores.

There is, however, a lack of research examining the process by which testing professionals can develop appropriate testing accommodations. Although several authors and test developers in their test manuals indicate the types of accommodations that may be appropriate or inappropriate with selected disabilities (e.g., Berg, Wacker, & Steege, 1995; Braden & Hannah, 1998; Bradley-Johnson, 1994; Reschly & Grimes, 1995), the practitioner is not presented with a process to use in making such determinations.

Prerequisites for Developing Accommodations

Examiner prerequisites for testing students with disabilities are knowledge of the disability and experience in working with individuals with that disability. Special education textbooks (e. g. Hallahan & Kauffman, 1997; Haring, McCormick, & Haring, 1994; Heward, 1996) as well as *Best Practices in School Psychology III* (Thomas & Grimes, 1995) are sources for the knowledge prerequisite. Equally important, however, is direct experience with the disability. It is essential that the examiner be familiar with the disability and feel comfortable in working with individuals with the disability. This type of experience is usually obtained during professional training but also can be gained by spending time in classrooms with students with disabilities, working with special education teachers and their students, and working with testing professionals who specialize in assessing students with disabilities, particularly low-incidence disabilities.

Likewise, when an accommodation is developed for an examinee, it is essential that the examinee feel comfortable with the accommodation and have direct experience with it. For example, allowing an individual to use a word processor instead of writing a response by hand would not be appropriate if the student has never used a word processor. In addition, the testing professional should be aware of any accommodations that may have been used in previous evaluations or are regularly used in the classroom. This information should be obtained prior to determining the need for testing accommodations. Finally, current information on the student's medical condition is important. These data may include functional visual assessments and hearing acuity results in the case of sensory impairments.

Developing Testing Accommodations

Testing accommodations should be developed only when no alternative measures exist. The following ten-step process can be used to guide decision making about what test to choose, what accommodations may be needed, and whether those accommodations will alter the construct being tested or the interpretation of results.

Step 1

The first step in developing testing accommodations is to determine the student's receptive skills. The examiner must determine whether the disability places limitations on the student's ability to understand visual or auditory material. Will the student be able to see the test materials, test questions, or any visual stimuli that are used? Will the student be able to hear the test directions or any verbal stimuli that are used? Any limitations in these receptive skills should be noted.

Step 2

The second step is to determine the student's expressive skills. The examiner must determine whether the disability places limitations on the student's ability to respond verbally or motorically to test items. Because many test items require a verbal response, the examiner must determine whether any limitations exist in this area. Some test items require motor responses, which may range from pointing to a response, to manipulating puzzle pieces and blocks, to copying marks or symbols with a pencil, to writing from one word to a sentence or paragraph or more. Does the student have the necessary physical skills to complete these tasks?

Step 3

The third step is to determine the construct, or specific skills, being measured. This is a crucial step because some test accommodations may have the effect of altering the construct being measured. The examiner needs to clearly determine what is to be measured so that an appropriate test can be used. An appropriate test is one that reliably and validly measures the skills the examiner has indicated and *does not require* expressive or receptive skills that the student lacks due to the disability.

Step 4

The fourth step in developing testing accommodations is to determine the purpose or purposes of assessment. Is it to make norm-based comparisons? Is it to determine whether the individual has mastered a particular skill or set of skills? Is it for program planning purposes? Is it for developing academic interventions? Is there a combination of purposes? This distinction is of utmost importance because the degree to which a test can be modified to accommodate individuals with disabilities and continue to produce valid scores is dependent, in part, on the purpose of the test.

In norm-referenced tests, comparisons are made between the individual's performance and the performance of individuals in the normative sample. The purpose of testing is to determine relative standing. The information being sought is how the student's performance compares with that of others of similar age, grade, background, etc. The emphasis is placed on whether the person is functioning above, below, or on par with similar individuals. Modifications in test stimuli, test procedures, or response format may reduce the meaningfulness of the test norms, as norm-referenced tests are based on the assumption that the same stimuli were administered in the same way to all students. Thus, normative comparisons under conditions of accommodation need to be interpreted very cautiously. The results could be used to determine whether the student possesses certain skills, such as being able to define specific vocabulary words. However, any normative comparisons would be inappropriate unless the norm group consists of similarly accommodated individuals.

Criterion-referenced tests, in contrast, are designed to determine level of skill development and whether the student possesses specific skills, rather than to make normative comparisons. Thus, accommodations in testing, although still important, do not have the same impact on the interpretation of scores as with norm-referenced tests.

Step 5

The fifth step in the process is to determine the test or tests to be used. This decision "must be based on the characteristics of the student . . . such as age, sensory status, language competencies, and acculturation" (Reschly & Grimes, 1995, p. 769). Best practice dictates that a standard or mandatory test best not be used. "Familiarity with a variety of instruments and knowledge of various disabling conditions are essential to choice of measures and interpretation of results" (Reschly & Grimes, 1995, p. 769).

Step 6

The sixth step involves a determination of the receptive skills and expressive skills required by the test or tests that have been selected. This step involves an analysis of how the test stimuli are presented (visually, verbally, or a combination of the two) and the response format of the test. How are students expected to express their responses? Many tests require verbal responses; others may require the manipulation of blocks or puzzles or a written response or pointing to the correct response or copying a design or symbols.

Step 7

In the seventh step, the examiner determines whether the student's receptive and expressive skills are sufficient for understanding the test items and responding appropriately. This determination is completed by comparing the answers to steps 1, 2, and 6.

Step 8

Once this analysis is completed, the examiner must use professional judgment to decide whether the set of skills needed for completing the test and the set of skills possessed by the student are sufficiently well matched to permit use of the test or tests. If they are, then testing can proceed. If not, the examiner must determine the type of accommodation that will be needed. The guiding principle in determining needed accommodations is that the accommodations should allow the student with disabilities to be assessed fairly and not be penalized as a result of the disability.

Step 9

In the ninth step, the examiner determines whether the necessary accommodations will compromise the test results. This decision rests heavily on the purpose of the assessment. If the purpose of assessment involves norm-based comparisons, several issues must be considered:

- Were individuals with disabilities included in the standardization sample? If so, were any of them provided with testing accommodations? If the answer to both these questions is yes, then the examiner can have greater confidence in making normative comparisons because the student would not have been specifically excluded from the standardization sample. If the answer to both these questions is no, or if accommodations were not made for individuals with disabilities in the standardization sample, then one must be more cautious in making normative comparisons.

- Have any specific accommodations been developed for the particular test? Consulting the test manual and contacting the publisher of the test are some ways to obtain this information.
- Does the testing modification alter the construct that is being measured? In other words, does the test measure the same construct with the accommodation as without? If the constructs being measured are not the same, then the accommodation is not appropriate. For example, a reading comprehension test that requires the individual to read a passage and verbally answer questions about it would be fundamentally altered by reading the passage to the student and having the student verbally answer questions about it. In this case the original construct, reading comprehension, is not being measured in the altered format; rather listening comprehension is being measured. Thus, the testing accommodation, although well intentioned, is not appropriate.

After answering these questions, the examiner must examine each proposed testing accommodation and determine whether the accommodation is appropriate to the purpose of the test and whether such an accommodation can be made. This step involves answering two questions. Does the accommodation alter the construct being measured by the test? Is the accommodation of sufficient magnitude that a comparison of scores between students with and without the accommodation is not appropriate? This decision should be made very carefully based on author and publisher recommendations, previous research, and finally, professional judgment.

If sufficient accommodations cannot be made, then the examiner must look for other ways to assess the skill or construct in question. In order to accomplish this, the examiner must be familiar with as many instruments as possible, as recommended by Reschly and Grimes (1995).

Step 10

In the final step the examiner carefully documents the accommodations necessary and describes any cautions or limitations in interpreting test results.

Following this procedure will help ensure that only the appropriate and necessary accommodations are made and that the test results are not compromised in the process.

References

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.

Berg, W. K., Wacker, D. P., & Steege, M. W. (1995). Best practices in assessment with persons who have severe or profound handicaps. In A. Thomas and J. Grimes (Eds.), *Best practices in school psychology III*. Washington, DC: National Association of School Psychologists.

Braden, J. P., & Hannah, J. M. (1998). Assessment of hearing-impaired and deaf children with the WISC-III. In A. Prifitera and D. Saklofske (Eds.), *WISC-III clinical use and interpretation*. San Diego: Academic Press.

Bradley-Johnson, S. (1994). *Psychoeducational assessment of students who are visually impaired or blind: Infancy through high school* (2nd ed.). Austin, TX: PRO-ED.

Hallahan, D. P., & Kauffman, J. M. (1997). *Exceptional learners: Introduction to special education* (7th ed.). Needham Heights, MA: Allyn & Bacon.

Haring, N. G., McCormick, L., & Haring, T. (1994). *Exceptional children and youth* (6th ed.). Columbus, OH: Merrill/Prentice Hall.

Heward, W. L. (1996). *Exceptional children: An introduction to special education* (5th ed.). Columbus, OH: Merrill/Prentice Hall.

Olson, J. F., & Goldstein, A. A. (1997). *The inclusion of students with disabilities and limited English proficient students in large-scale assessment: A summary of recent progress*. Washington, DC: U.S. Department of Education.

Reschly, D. J., & Grimes, J. P. (1995). Best practices in intellectual assessment. In A. Thomas and J. Grimes (Eds.), *Best practices in school psychology III*. Washington, DC: National Association of School Psychologists.

Thomas, A., & Grimes, J. (Eds.). (1995). *Best practices in school psychology III*. Washington, DC: National Association of School Psychologists.

Thurlow, M. L., Elliott, J. L., & Ysseldyke, J. E. (1998). *Testing students with disabilities: Practical strategies for complying with district and state requirements*. Thousand Oaks, CA: Corwin Press.

About the Author

Dougles K. Smith is currently director of programs in school psychology at the University of Albany—State University of New York. He obtained his Ph.D., Ed.S., and M.Ed. degrees in school psychology from Georgia State University. Current research interests include psychoeducational assessment issues in general and developing individual testing accommodations for students with disabilities in particular. Smith is author of *Essentials of Individual Achievement Assessment* (2001) and co-editor of the forthcoming *Assessing People With Disabilities in Educational, Employment, and Counseling Settings*, as well as numerous journal and chapter articles. Smith was named Outstanding Faculty Member of 1987 in the College of Education at the University of Wisconsin–River Falls.

1. An expanded version of this paper is scheduled to appear in Ekstrom, R. B., & Smith, D. (Eds.). *The assessment of individuals with disabilities: A guide for practitioners*. Washington, DC: American Psychological Association.

Special Assessment Topics

Development of a Statement of Test Takers' Rights and Responsibilities
Kurt F. Geisinger

Counseling Outcome Research: Making Practical Choices for Real-World Applications
Darcy H. Granello and Paul F. Granello

Communicating Assessment Results in the Counseling Interview
Albert B. Hood

Response to Surveys by Electronic and Traditional Means
David J. Lundberg

Using the Internet to Enhance Test Interpretation
James P. Sampson Jr.

Use of the Kaufman Adolescent and Adult Intelligence Test (KAIT) in the New Millennium
Douglas K. Smith

Writing Multiple-Choice Test Items
Nicholas A. Vacc, Larry C. Loesch, and Ruth E. Lubik

Chapter Thirteen

Development of a Statement of Test Takers' Rights and Responsibilities

Kurt F. Geisinger[1]

Abstract

Various professional organizations whose members are involved in testing have acknowledged the rights of test takers in their standards. The Joint Committee on Testing Practices, composed of delegates from six professional associations, has produced an enumeration of rights and responsibilities of test takers. The development of this draft document is summarized and the document itself is included in a chapter appendix.

In this article, I would like to describe the process that has led to the development of a draft document enumerating test takers' rights and responsibilities from procedural and historical perspectives. For four years (1993–1997), I represented the American Psychological Association on the Joint Committee on Testing Practices (JCTP). The JCPT is the embodiment of a relatively rare form of interdisciplinary dialogue. It is composed of delegates from six professional associations: American Counseling Association, American Educational Research Association,[2] American Psychological Association, the American Speech-Language-Hearing Association, the National Association of School Psychologists, and the National Council on Measurement in Education.[3] The JCTP divides into working groups to tackle thorny testing problems that fall under the general theme of test use. Among the products that have been developed in the past 10 years of JCTP activities are the Code of Fair Testing Practices in Education (JCTP, 1988) in English and Spanish versions, a variety of reports on test misuse; a volume entitled *Responsible Test Use* (Eyde et al., 1993);

and a videotape entitled *The ABCs of School Testing* (1993), which is accompanied by an instructional manual. We have two primary current projects, one of which is exploring the most effective ways of testing individuals with disabilities and the second of which is the subject of this article. We have developed a draft statement of test takers' rights and responsibilities[4] and we are thoroughly committed to the idea of listening to the feedback from the public and the professions that we represent and making whatever modifications are appropriate based upon recommendations coming from this forum and others in the future. I will now provide a synopsis of the historical development of this draft document (see the appendix for a copy of the document).

A Little History

In many cases, testing is a public practice, and as such is controlled, regulated, and influenced by a variety of sources. Some of these agencies include the government (e.g., the *Uniform Guidelines on Employee Selection Procedures,* Equal Employment Opportunity Commission et al., 1978), professional associations (e.g., *Standards for Educational and Psychological Testing,* AERA, APA, & NCME, 1985/1999), and organizations themselves (e.g., *ETS Principles, Policies, and Procedural Guidelines Regarding ETS Products and Services,* Educational Testing Service, 1979).

Mel Novick (1981) summarized roles of governmental agencies, professional standards, and state and federal guidelines on the profession and practice of testing. Novick rightly acknowledged that there are three participants in the testing process: the test producer (who "develops, markets, and/or administers and scores the test" (p. 1035), the test user (who chooses to use a given measure for a specified purpose), and the test taker (who, with a greater or lesser degree of choice, completes the measure) under conditions set by the test producer, the test user, or some combination of the two.

The *Standards for Educational and Psychological Testing* (AERA et al., 1985) first acknowledged the rights of test takers in its 1985 emanation. The last chapter of that edition of the technical standards was entitled "Protecting the Rights of Test Takers," and earlier in the document other chapters were devoted to the special problems inherent in testing linguistic minorities and people with handicapping conditions. Thus in 1985 the profession clearly began the process of acknowledging the importance of test takers as participants in the testing process. It might be noted that the discussion of test takers' rights has typically been related to the delivery of specific kinds of information. It might be held, for example, that individuals have a right to valid and fair assessments. However, such pronouncements would be seen as relating to test validation and fairness rather than test takers' rights per se.

The technical standards present 10 general rights that individuals who take tests are deemed to have. These rights might be broken into two types: (a) test score reporting and interpretation, including the access to such information; and (b) test score cancellation and the processes used to make such decisions. The former area encompasses test takers' rights relating to (a) the circumstances under which informed consent is required prior to actual testing; (b) the school, clinical, and counseling applications in which test takers or legal representatives should receive transmittal of and an explanation about test results; (c) the principle that the names of individual test takers generally should not be associated with their performance in public pronouncements related to the testing; (d) the maintenance of reasonable security with respect to databases and the like; and (e) the avoidance stigmatizing labels wherever possible.

The second area of test takers' rights concerns test score cancellation and the processes used by test users and test publishers to make such judgments. Such cancellations typically occur due to suspected misconduct. In general, the standards hold that (a) the procedures used to make such decisions should be explained to a test taker facing the possibility of score cancellation; (b) the test taker should be notified that the investigative process is, in fact, ongoing; (c) in certain prescribed circumstances (primarily where tests are used in educational admissions, licensing, and certification) the test taker should be allowed to provide evidence to be considered as part of the entire score cancellation investigation; and (d) all available evidence should be reviewed in educational admissions, licensing, and certification circumstances.

The preceding lists briefly summarize the rights provided to those taking tests under the technical standards for testing that are currently in force. The *Code of Fair Testing Practices in Education* (JCTP, 1988) provides similar informational rights to test takers although, being a shorter document, it provides less specificity. The code suggests that test developers and users should "provide test takers" with "the information they need to be familiar with the coverage of the test, the types of question formats, the directions, and appropriate test-taking strategies" and, for optional tests, they should provide information to possible test takers so that they can be best informed and able to decide whether or not to take the examination (p. 4). The code also states that test developers and test users should inform test takers and their parents or guardians of any rights that they may have, of the procedures that they may use to register complaints about the testing or to have problems resolved, and the nature and security of the test scores after the actual testing.

At a 1991 symposium sponsored by the JCTP at the annual meeting of the American Psychological Association, Robert Perloff and I considered the circumstances under which informational feedback to those taking assessments was required and preferred. Our goal at the time, one that continues into the present, was and is to initiate a dialogue among participants in the testing process to advance our usefulness to the institutions and individuals that we serve. Tensions between testers and test takers remain high, as evidenced by the frequency with which testing issues receive attention in the press. The testing profession needs to converse meaningfully with those who take our assessments.

Dr. Perloff was concerned, for example, that college students and others completing tests, surveys, and other measures were becoming disaffected and unwilling to continue completing such instruments. He believed that the reason for such recalcitrant behavior was that the test takers were tired of performing without receiving feedback regarding the nature of their responses. He stated, "The responsibility, I submit, is ours, the testers, the publishers, the test authors, the professional testing community, to make tests that we are willing to go to the mat for. There is no, none whatever, medical test conducted on a patient whose results the patient does not learn about, in specific units or numbers on a scale, and interpreted by the patient's physician or other professional" (Perloff, 1991, pp. 2–3). He summarized his perspective again with the following goal: "The aim that I am (perhaps naively) championing is that all test scores—good, bad, or indifferent—be disclosed to the examinee, unless he or she explicitly prefers not to know; ignorance, of course, is the person's right" (p. 3). He stated the hope that by the end of the 1990s all results of all types of assessments would be provided to the individual assessed in an understandable, reasonable fashion. Perloff continued that if we are to contend that our measures are valid and meaningful, then we should forthrightly indicate that we are not ashamed of our measures and, in fact, should share the results of our assessment openly with our clientele. He called as well for informational brochures that could be provided to test takers at the time they receive the results of their assessments, to help them interpret the meaning of their results.

My response to Dr. Perloff in that setting (Geisinger, 1991) primarily set boundaries on the release of information. I was and continue to be concerned, for example, about the release of test score information taken from measures that have not been validated, and the costs and expertise required to provide feedback in some settings. I might argue that it is not even possible to interpret an individual's performance on a measure that has not yet been validated. I value the notion of providing informative pamphlets to test score recipients so that they can interpret their own performance with well-written, concise guidance. There is, however, information that one probably should not learn without a knowledgeable professional present and without the opportunity to ask pertinent questions. The costs of such individual feedback would be prohibitive in many settings and might remove the utility from the testing process. I also reported concern about the provision of feedback related to performance on tests the purpose of which is to predict future behavior rather than to represent a domain of present behavior. In some cases, especially in some clinical and industrial testing settings, the predictive measures have little meaning in and of themselves. I concluded my paper with the following: "Validation research is needed for many of the tests that we use to help us develop interpretations that we can share in a meaningful way with test takers. More critically, at this stage, protracted dialogue should occur among the various parties—test takers, test makers, test users—to define what the nature of feedback should be. Perhaps a 'Bill of Test Takers' Rights' is needed. I would urge the development of such a document" (p. 9).

The JCTP decided in 1993–94 to study test takers' rights and perhaps to develop an enumeration of the rights we believed that test takers have or should have based on a literature review. All of our constituent groups supported this initial proposal. Until about 1995, a small but dedicated group of committee members followed the charge put forth to initiate dialogue regarding both the provision of feedback to test takers and, more generally, the rights of test takers. As part of the development of our document, we were committed to public discussion about the rights of test takers. In fact, among the earliest feedback we received from other professionals was that we needed to consider the responsibilities of test takers as well as their rights. The current (and final) document includes both of these components. I note that, following our lead, the 1999 *Standards for Educational and Psychological Testing* also have combined these two related issues. The symposium held at the Assessment '98 convention was one of many fora where we solicited the advice and counsel of professionals before final revisions to the draft document were made. Even now that the document is finished (Test Taker Rights and Responsibilities Working Group of the JCTP, 2000), we know that it continues to be formative and will be refined over time. Nevertheless, we are optimistic and hope that in its current version it will serve to improve the ability of the testing profession to advance our society.

References

American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). (1985/1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Educational Testing Service. (1979). *ETS principles, policies, and procedural guidelines regarding ETS products and services*. Princeton, NJ: Author.

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, and Department of Justice. (1978). Adoption by four agencies of Uniform Guidelines on Employee Selection Procedures. *Federal Register, 43*, 38290–38315.

Eyde, L. D., Robertson, G. J., King, S. E., Moreland, K. L., Robertson, A. G., Shewan, C. M., Harrison, P. L., Porch, B. E., Hammer, A. L., & Primoff, E. S. (1993). *Responsible test use: Case studies for assessing human behavior*. Washington, DC: American Psychological Association.

Geisinger, K. F. (1991, April). Disclosing interpreted test scores to test takers: What are the problems? Paper presented to the symposium Understanding Test Results—What Users and Examinees Should Know (Jo-Ida C. Hansen, Chair) at the annual meeting of the American Psychological Association, San Francisco, CA.

JCTP. (1993, August). *The ABCs of school testing*. Washington, DC: National Council on Measurement in Education.

JCTP. (1988). *Code of fair testing practices in education*. Washington, DC: Author.

Novick, M. R. (1981). Federal guidelines and professional standards. *American Psychologist, 36*, 1035–1046.

Perloff, R. (1991, April). The case for disclosing interpreted test scores to test examinees. Paper presented to the symposium Understanding Test Results—What Users and Examinees Should Know (Jo-Ida C. Hansen, Chair) at the annual meeting of the American Psychological Association, San Francisco, CA.

Test Taker Rights and Responsibilities Working Group of the Joint Committee on Testing Practices. (2000). *Rights and responsibilities of test takers: Guidelines and expectations.* Washington, DC: American Psychological Association.

About the Author

Kurt F. Geisinger is vice president for academic affairs and professor of psychology at the University of St. Thomas in Houston, Texas. He is co-author of *Test Interpretation for Members of Diverse Groups* and author of *Psychological Testing of Hispanics*, both published by the American Psychological Association. He has been a member of the American Psychological Association Committee on Psychological Testing and Assessment, the Middle States Regional Council for the College Board, the College Board SAT Committee, and the chair of the Technical Advisory Committee for the Graduate Record Examination at Educational Testing Service. From 1993 to 1997 he co-chaired the Joint Committee on Testing Practices, a committee composed of representatives from five professional associations, and he initiated and co-chaired the working group that developed the Statement of Test Takers' Rights and Responsibilities, for which he was awarded the Distinguished Contributions award by the Association for Assessment in Counseling in 2001. He presently serves as a consulting editor for the *College Board Review*, the *International Journal of Testing*, and *Educational Research Quarterly*. A Fellow of APA, his specialties include psychometrics, test construction, the testing of language minorities and those with disabilities, and the translation of tests.

1. This article was originally presented at the Assessment '98: Assessment for Change—Changes in Assessment conference, St. Petersburg, FL, January 16–18, 1998.

2. The American Educational Research Association rejoined the JCTP near the end of 1997. It had previously been a member, but not during the development of the Test Takers' Rights and Responsibilities document.

3. In 2000 the National Association of Test Directors also joined the JCTP.

4. This document is reprinted in the appendix and is currently

available at the JCTP website, which may be found within that of the American Psychological Association at http://www.apa.org/science/jctpweb.html.

Appendix: Rights and Responsibilities of Test Takers: Guidelines and Expectations

Preamble

The intent of this statement is to enumerate and clarify the expectations that test takers may reasonably have about the testing process, and the expectations that those who develop, administer, and use tests may have of test takers. Tests are defined broadly here as psychological and educational instruments developed and used by testing professionals in organizations such as schools, industries, clinical practice, counseling settings and human service, and other agencies, including those assessment procedures and devices that are used for making inferences about people in the above-named settings. The purpose of the statement is to inform and to help educate not only test takers, but also others involved in the testing enterprise so that measurements may be most validly and appropriately used. This document is intended as an effort to inspire improvements in the testing process and does not have the force of law. Its orientation is to encourage positive and high-quality interactions between testing professionals and test takers.

The rights and responsibilities listed in this document are neither legally based nor inalienable rights and responsibilities such as those listed in the United States of America's Bill of Rights. Rather, they represent the best judgments of testing professionals about the reasonable expectations that those involved in the testing enterprise (test producers, test users, and test takers) should have of each other. Testing professionals include developers of assessment products and services, those who market and sell them, persons who select them, test administrators and scorers, those who interpret test results, and trained users of the information. Persons who engage in each of these activities have significant responsibilities that are described elsewhere, in documents such as those that follow (American Association for Counseling and Development, 1988; American Speech-Language-Hearing Association, 1994; Joint Committee on Testing Practices, 1988; National Association of School Psychologists, 1992; National Council on Measurement in Education, 1995).

In some circumstances, the test developer and the test user may not be the same person, group of persons, or organization. In such situations, the professionals involved in the testing should clarify, for the test taker as well as for themselves, who is responsible for each aspect of the testing process. For example, when an individual chooses to take a college admissions test, at least three parties are involved in addition to the test taker: the test developer and publisher, the individuals who administer the test to the test taker, and the institutions of higher education who will eventually use the information. In such cases a test taker may need to request clarifications about their rights and responsibilities. When test takers are young children (e.g., those taking standardized tests in the schools) or are persons who spend some or all their time in institutions or are incapacitated, parents or guardians may be granted some of the rights and responsibilities, rather than, or in addition to, the individual.

Perhaps the most fundamental right test takers have is to be able to take tests that meet high professional standards, such as those described in *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) as well as those of other appropriate professional associations. This statement should be used as an adjunct, or supplement, to those standards. State and federal laws, of course, supersede any rights and responsibilities that are stated here.

References

American Association for Counseling and Development (now American Counseling Association) & Association for Measurement and Evaluation in Counseling and Development (now Association for Assessment in Counseling). (1989). *Responsibilities of users of standardized tests: RUST statement revised.* Alexandria, VA: Author.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

American Speech-Language-Hearing Association. (1994). Protection of rights of people receiving audiology or speech-language pathology services. *ASHA 36*, 60–63.

Eyde, L. D., Robertson, G. J., King, S. E., Moreland, K. L., Robertson, A. G., Shewan, C. M., Harrison, P. L., Porch, B. E., Hammer, A. L., & Primoff, E. S. (1993). *Responsible test use: Case studies for assessing human behavior.* Washington, DC: American Psychological Association.

Joint Committee on Testing Practices. (1993). *ABCs of School Testing.* [Video] Washington, DC: National Council on Measurement in Education.

Joint Committee on Testing Practices. (1988). *Code of fair testing practices in education.* Washington, DC: American Psychological Association.

National Association of School Psychologists. (1992). *Standards for the provision of school psychological services.* Silver Springs, MD: Author.

National Council on Measurement in Education. (1995). *Code of professional responsibilities in educational measurement.* Washington, DC: Author.

It is recommended that the following guidelines for test takers be widely circulated.

The Rights and Responsibilities of Test Takers: Guidelines and Expectations

Test Taker Rights and Responsibilities Working Group
of the Joint Committee on Testing Practices
August 1998

As a test taker, you have the right to:

1. Be informed of your rights and responsibilities as a test taker.
2. Be treated with courtesy, respect, and impartiality, regardless of your age, disability, ethnicity, gender, national origin, religion, sexual orientation, or other personal characteristics.
3. Be tested with measures that meet professional standards and that are appropriate, given the manner in which the test results will be used.
4. Receive a brief oral or written explanation prior to testing about the purpose(s) for testing, the kind(s) of tests to be used, if the results will be reported to you or to others, and the planned use(s) of the results. If you have a disability, you have the right to inquire and receive information about testing accommodations. If you have difficulty in comprehending the language of the test, you have a right to know in advance of testing whether any accommodations may be available to you.
5. Know in advance of testing when the test will be administered, if and when test results will be available to you, and if there is a

fee for testing services that you are expected to pay.

6. Have your test administered and your test results interpreted by appropriately trained individuals who follow professional codes of ethics.
7. Know if a test is optional and learn of the consequences of taking or not taking the test, fully completing the test, or canceling the scores. You may need to ask questions to learn these consequences.
8. Receive a written or oral explanation of your test results within a reasonable amount of time after testing and in commonly understood terms.
9. Have your test results kept confidential to the extent allowed by law.
10. Present concerns about the testing process or your results and receive information about procedures that will be used to address such concerns.

As a test taker, you have the responsibility to:

1. Read and/or listen to your rights and responsibilities as a test taker.
2. Treat others with courtesy and respect during the testing process.
3. Ask questions prior to testing if you are uncertain about why the test is being given, how it will be given, what you will be asked to do, and what will be done with the results.
4. Read or listen to descriptive information in advance of testing and listen carefully to all test instructions. You should inform an examiner in advance of testing if you wish to receive a testing accommodation or if you have a physical condition or illness that may interfere with your performance on the test. If you have difficulty comprehending the language of the test, it is your responsibility to inform an examiner.
5. Know when and where the test will be given, pay for the test if required, appear on time with any required materials, and be ready to be tested.
6. Follow the test instructions you are given and represent yourself honestly during the testing.
7. Be familiar with and accept the consequences of not taking the test, should you choose not to take the test.
8. Inform appropriate person(s), as specified to you by the organization responsible for testing, if you believe that testing conditions affected your results.
9. Ask about the confidentiality of your test results, if this aspect concerns you.

10. Present concerns about the testing process or results in a timely, respectful way, if you have any.

The Rights of Test Takers: Guidelines for Testing Professionals

Test takers have the rights described below. It is the responsibility of the professionals involved in the testing process to ensure that test takers receive these rights.

1. Because test takers have the right to be informed of their rights and responsibilities as test takers, it is normally the responsibility of the individual who administers a test (or the organization that prepared the test) to inform test takers of these rights and responsibilities.
2. Because test takers have the right to be treated with courtesy, respect, and impartiality, regardless of their age, disability, ethnicity, gender, national origin, race, religion, sexual orientation, or other personal characteristics, testing professionals should:
    a. Make test takers aware of any materials that are available to assist them in test preparation. These materials should be clearly described in test registration and/or test familiarization materials.
    b. See that test takers are provided with reasonable access to testing services.
3. Because test takers have the right to be tested with measures that meet professional standards that are appropriate for the test use and the test taker, given the manner in which the results will be used, testing professionals should:
    a. Take steps to utilize measures that meet professional standards and are reliable, relevant, useful given the intended purpose and are fair for test takers from varying societal groups.
    b. Advise test takers that they are entitled to request reasonable accommodations in test administration that are likely to increase the validity of their test scores if they have a disability recognized under the Americans with Disabilities Act or other relevant legislation.
4. Because test takers have the right to be informed, prior to testing, about the test's purposes, the nature of the test, whether test results will be reported to the test takers, and the planned use of the results (when not in conflict with the testing purposes), testing professionals should:
    a. Give or provide test takers with access to a brief description about the test purpose (e.g., diagnosis, placement, selection,

etc.) and the kind(s) of tests and formats that will be used (e.g., individual/group, multiple-choice/free response/performance, timed/untimed, etc.), unless such information might be detrimental to the objectives of the test.

b. Tell test takers, prior to testing, about the planned use(s) of the test results. Upon request, the test taker should be given information about how long such test scores are typically kept on file and remain available.

c. Provide test takers, if requested, with information about any preventative measures that have been instituted to safeguard the accuracy of test scores. Such information would include any quality control procedures that are employed and some of the steps taken to prevent dishonesty in test performance.

d. Inform test takers, in advance of the testing, about required materials that must be brought to the test site (e.g., pencil, paper) and about any rules that allow or prohibit use of other materials (e.g., calculators).

e. Provide test takers, upon request, with general information about the appropriateness of the test for its intended purpose, to the extent that such information does not involve the release of proprietary information. (For example, the test taker might be told, "Scores on this test are useful in predicting how successful people will be in this kind of work" or "Scores on this test, along with other information, help us to determine if students are likely to benefit from this program.")

f. Provide test takers, upon request, with information about re-testing, including if it is possible to re-take the test or another version of it, and if so, how often, how soon, and under what conditions.

g. Provide test takers, upon request, with information about how the test will be scored and in what detail. On multiple-choice tests, this information might include suggestions for test taking and about the use of a correction for guessing. On tests scored using professional judgment (e.g., essay tests or projective techniques), a general description of the scoring procedures might be provided except when such information is proprietary or would tend to influence test performance inappropriately.

h. Inform test takers about the type of feedback and interpretation that is routinely provided, as well as what is available for a fee. Test takers have the right to request and receive information regarding whether or not they can obtain copies of their test answer sheets or their test materials, if they can have their scores verified, and if they may cancel their test results.

i. Provide test takers, prior to testing, either in the written instructions, in other written documents or orally, with answers to questions that test takers may have about basic test administration procedures.

j. Inform test takers, prior to testing, if questions from test takers will not be permitted during the testing process.

k. Provide test takers with information about the use of computers, calculators, or other equipment, if any, used in the testing and give them an opportunity to practice using such equipment, unless its unpracticed use is part of the test purpose, or practice would compromise the validity of the results, and to provide a testing accommodation for the use of such equipment, if needed.

l. Inform test takers that, if they have a disability, they have the right to request and receive accommodations or modifications in accordance with the provisions of the Americans with Disabilities Act and other relevant legislation.

m. Provide test takers with information that will be of use in making decisions if test takers have options regarding which tests, test forms or test formats to take.

5. Because that test takers have a right to be informed in advance when the test will be administered, if and when test results will be available, and if there is a fee for testing services that the test takers are expected to pay, test professionals should:
   a. Notify test takers of the alteration in a timely manner if a previously announced testing schedule changes, provide a reasonable explanation for the change, and inform test takers of the new schedule. If there is a change, reasonable alternatives to the original schedule should be provided.
   b. Inform test takers prior to testing about any anticipated fee for the testing process, as well as the fees associated with each component of the process, if the components can be separated.
6. Because test takers have the right to have their tests administered and interpreted by appropriately trained individuals, testing professionals should:
   a. Know how to select the appropriate test for the intended purposes.

   b. When testing persons with documented disabilities and other special characteristics that require special testing conditions and/or interpretation of results, have the skills and knowledge for such testing and interpretation.
   c. Provide reasonable information regarding their qualifications, upon request.
   d. Insure that test conditions, especially if unusual, do not unduly interfere with test performance. Test conditions will normally be similar to those used to standardize the test.
   e. Provide candidates with a reasonable amount of time to complete the test, unless a test has a time limit.
   f. Take reasonable actions to safeguard against fraudulent actions (e.g., cheating) that could place honest test takers at a disadvantage.
7. Because test takers have the right to be informed about why they are being asked to take particular tests, if a test is optional, and what the consequences are should they choose not to complete the test, testing professionals should:
   a. Normally only engage in testing activities with test takers after the test takers have provided their informed consent to take a test, except when testing without consent has been mandated by law or governmental regulation, or when consent is implied by an action the test takers have already taken (e.g., such as when applying for employment and a personnel examination is mandated).
   b. Explain to test takers why they should consider taking voluntary tests.
   c. Explain, if a test taker refuses to take or complete a voluntary test, either orally or in writing, what the negative consequences may be to them for their decision to do so.
   d. Promptly inform the test taker if a testing professional decides that there is a need to deviate from the testing services to which the test taker initially agreed (e.g., should the testing professional believe it would be wise to administer an additional test or an alternative test), and provide an explanation for the change.
8. Because test takers have a right to receive a written or oral explanation of their test results within a reasonable amount of time after testing and in commonly understood terms, testing professionals should:
   a. Interpret test results in light of one or more additional considerations (e.g., disability, language proficiency), if those considerations are relevant to the purposes of the

test and performance on the test, and are in accordance with current laws.

b. Provide, upon request, information to test takers about the sources used in interpreting their test results, including technical manuals, technical reports, norms, and a description of the comparison group, or additional information about the test taker(s).

c. Provide, upon request, recommendations to test takers about how they could improve their performance on the test, should they choose or be required to take the test again.

d. Provide, upon request, information to test takers about their options for obtaining a second interpretation of their results. Test takers may select an appropriately trained professional to provide this second opinion.

e. Provide test takers with the criteria used to determine a passing score, when individual test scores are reported and related to a pass-fail standard.

f. Inform test takers, upon request, how much their scores might change, should they elect to take the test again. Such information would include variation in test performance due to measurement error (e.g., the appropriate standard errors of measurement) and changes in performance over time with or without intervention (e.g., additional training or treatment).

g. Communicate test results to test takers in an appropriate and sensitive manner, without use of negative labels or comments likely to inflame or stigmatize the test taker.

h. Provide corrected test scores to test takers as rapidly as possible, should an error occur in the processing or reporting of scores. The length of time is often dictated by individuals responsible for processing or reporting the scores, rather than the individuals responsible for testing, should the two parties indeed differ.

i. Correct any errors as rapidly as possible if there are errors in the process of developing scores.

9. Because test takers have the right to have the results of tests kept confidential to the extent allowed by law, testing professionals should:

   a. Insure that records of test results (in paper or electronic form) are safeguarded and maintained so that only individuals who have a legitimate right to access them will be able to do so.

   b. Should provide test takers, upon request, with information regarding who has a legitimate right to access their test results (when individually identified) and in what form. Testing professionals should respond appropriately to questions regarding the reasons why such individuals may have access to test results and how they may use the results.

c. Advise test takers that they are entitled to limit access to their results (when individually identified) to those persons or institutions, and for those purposes, revealed to them prior to testing. Exceptions may occur when test takers, or their guardians, consent to release the test results to others or when testing professionals are authorized by law to release test results.

d. Keep confidential any requests for testing accommodations and the documentation supporting the request.

10. Because test takers have the right to present concerns about the testing process and to receive information about procedures that will be used to address such concerns, testing professionals should:

    a. Inform test takers how they can question the results of the testing if they do not believe that the test was administered properly or scored correctly, or other such concerns.

    b. Inform test takers of the procedures for appealing decisions that they believe are based in whole or in part on erroneous test results.

    c. Inform test takers, if their test results are under investigation and may be canceled, invalidated, or not released for normal use. In such an event, that investigation should be performed in a timely manner. The investigation should use all available information that addresses the reason(s) for the investigation, and the test taker should also be informed of the information that he/she may need to provide to assist with the investigation.

    d. Inform the test taker, if that test taker's test results are canceled or not released for normal use, why that action was taken. The test taker is entitled to request and receive information on the types of evidence and procedures that have been used to make that determination.

The Responsibilities of Test Takers: Guidelines for Testing Professionals

Testing Professionals should take steps to ensure that test takers know that they have specific responsibilities in addition to their rights described above.

1. Testing professionals need to inform test takers that they should listen to and/or read their rights and responsibilities as a test taker and ask questions about issues they do not understand.
2. Testing professionals should take steps, as appropriate, to ensure that test takers know that they:
    a. Are responsible for their behavior throughout the entire testing process.
    b. Should not interfere with the rights of others involved in the testing process.
    c. Should not compromise the integrity of the test and its interpretation in any manner.
3. Testing professionals should remind test takers that it is their responsibility to ask questions prior to testing if they are uncertain about why the test is being given, how it will be given, what they will be asked to do, and what will be done with the results. Testing professionals should:
    a. Advise test takers that it is their responsibility to review materials supplied by test publishers and others as part of the testing process and to ask questions about areas that they feel they should understand better prior to the start of testing.
    b. Inform test takers that it is their responsibility to request more information if they are not satisfied with what they know about how their test results will be used and what will be done with them.
4. Testing professionals should inform test takers that it is their responsibility to read descriptive material they receive in advance of a test and to listen carefully to test instructions. Testing professionals should inform test takers that it is their responsibility to inform an examiner in advance of testing if they wish to receive a testing accommodation or if they have a physical condition or illness that may interfere with their performance. Testing professionals should inform test takers that it is their responsibility to inform an examiner if they have difficulty comprehending the language in which the test is given. Testing professionals should:
    a. Inform test takers that, if they need special testing arrangements, it is their responsibility to request appropriate accommodations and to provide any requested documentation as far in advance of the testing date as possible. Testing professionals should inform test takers about the documentation needed to receive a requested testing accommodation.
    b. Inform test takers that, if they request but do not receive

a testing accommodation, they could request information about why their request was denied.

5. Testing professionals should inform test takers when and where the test will be given, and whether payment for the testing is required. Having been so informed, it is the responsibility of the test taker to appear on time with any required materials, pay for testing services and be ready to be tested. Testing professionals should:
    a. Inform test takers that they are responsible for familiarizing themselves with the appropriate materials needed for testing and for requesting information about these materials, if needed.
    b. Inform the test taker, if the testing situation requires that test takers bring materials (e.g., personal identification, pencils, calculators, etc.) to the testing site, of this responsibility to do so.
6. Testing professionals should advise test takers, prior to testing, that it is their responsibility to:
    a. Listen to and/or read the directions given to them.
    b. Follow instructions given by testing professionals.
    c. Complete the test as directed.
    d. Perform to the best of their ability if they want their score to be a reflection of their best effort.
    e. Behave honestly (e.g., not cheating or assisting others who cheat).
7. Testing professionals should inform test takers about the consequences of not taking a test, should they choose not to take the test. Once so informed, it is the responsibility of the test taker to accept such consequences, and the testing professional should so inform the test takers. If test takers have questions regarding these consequences, it is their responsibility to ask questions of the testing professional, and the testing professional should so inform the test takers.
8. Testing professionals should inform test takers that it is their responsibility to notify appropriate persons, as specified by the testing organization, if they do not understand their results, or if they believe that testing conditions affected the results. Testing professionals should:
    a. Provide information to test takers, upon request, about appropriate procedures for questioning or canceling their test scores or results, if relevant to the purposes of testing.
    b. Provide to test takers, upon request, the procedures for reviewing, retesting, or canceling their scores or test

    results, if they believe that testing conditions affected their results and if relevant to the purposes of testing.
    c. Provide documentation to the test taker about known testing conditions that might have affected the results of the testing, if relevant to the purposes of testing.

9. Testing professionals should advise test takers that it is their responsibility to ask questions about the confidentiality of their test results, if this aspect concerns them.
10. Testing professionals should advise test takers that it is their responsibility to present concerns about the testing process in a timely, respectful manner.

Members of the JCTP Working Group on Test Taker Rights and Responsibilities
Kurt F. Geisinger, Ph.D. (Co-chair)
William Schafer, Ph.D. (Co-chair)
Gwyneth Boodoo, Ph.D.
Ruth Ekstrom, Ed.D.
Tom Fitzgibbon, Ph.D.
John Fremer, Ph.D.
Joanne Lenke, Ph.D.
Sharon Goldsmith, Ph.D.
Julie Noble, Ph.D.
Douglas Smith, Ph.D.
Nicholas Vacc, Ed.D.
Janet Wall, Ed.D.
Staff liaisons: Heather Fox, Ph.D., and Lara Frumkin, Ph.D.

Chapter Fourteen

Counseling Outcome Research: Making Practical Choices for Real-World Applications

Darcy H. Granello
Paul F. Granello

Abstract

Mental health practitioners are increasingly being called upon to demonstrate the effectiveness of their clinical interventions. Effectiveness studies are a type of outcome research that can provide useful information to clinicians and to managed care organizations.

In an age of managed care, counselors are increasingly being called upon to demonstrate the effectiveness of their clinical interventions (Granello, Granello, & Lee, 1999). The ability to demonstrate treatment success is rapidly becoming the standard by which reimbursement is judged (Sexton, 1996). In spite of these pressures, many counselors have been left unprepared to meet this new standard. Historically, mental health practitioners used professional judgment and theoretical beliefs to determine treatment interventions. Fee-for-service policies and insurance reimbursement were assumed, and insurance companies rarely questioned treatment decisions (Plante, Couchman, & Diaz, 1995). In the current practice environment, however, counselors who cannot demonstrate their successes may find themselves unable to survive professionally (Burlingame, Lambert, & Reisinger, 1995).

Although the demonstration of treatment effectiveness is increasing in importance, many mental health professionals and agencies have resisted participation in outcome measures, and there is widespread resistance among mental health professionals to beginning their own assessment programs (Plante, et al. 1995). Studies have revealed that the vast majority of mental health practitioners report that they do not read research or engage in research and believe that research has little or no impact on their counseling practices (Cohen, Sargent, & Sechrest, 1986; Falvey, 1989). In 1983, Norcross and Prochaska found that when presented with 14 reasons to select a particular approach or orientation with a client, the psychologists in their study rated outcome research 10th, just above "family experiences" and "own therapist's orientation." More recently, Norcross (2000) noted there was little evidence that this ranking had improved significantly during the past 17 years, although he predicted that the recent emphasis on the importance of outcome research should result in increased reliance on such research in the future. A recent survey found that although the majority of the clinical diplomates of the American Board of Professional Psychology (65%) supported the development of empirically supported treatments, the majority of respondents (54%) did not routinely use them in their practices (Plante, Anderson, & Boccaccini, 1999).

Both philosophical and practical concerns have been identified at the root of the resistance to engaging in outcome research and incorporating research results into practice. Philosophically, some providers have argued that the invasion of accountability into mental health care has negatively affected therapeutic decision making (Sherman, 1992). Some argue that the therapeutic process itself is not quantifiable (Mirin & Namerow, 1991) or that clinical flexibility, clinical judgment, and creative expression of theory should be valued more than scientific method and statistical analysis (Havens, 1994). Still others argue that time spent in evaluation could be better used in treatment (Plante, et al. 1995). Even among clinicians who are willing to conduct outcome research, practical concerns often stand in the way. Practitioners may erroneously believe that the task will be overwhelming or that a program of research will necessarily be costly, complex, and time-consuming (Granello et al., 1999). What has become apparent is that few mental health practitioners have received the training they need to conduct such research. Research methods courses in university programs often focus on understanding laboratory research with true experimental designs that are often impossible to implement in real-world assessment (Sandell, Blomberg, & Lazar, 1997). Thus, practitioners may be ill prepared to conduct their own outcome research, regardless of their willingness to do so.

The incorporation of already published outcome data into clinical practice plays a significant role in determining appropriate treatment interventions and the efficacy of various modalities (Sexton, 2000). Bridging the gap between research and practice is essential (Whiston & Coker, 2000). However, if a practitioner is willing to conduct his or her own outcome research, in conjunction with already published research to support general clinical interventions, the result will be enhanced quality of care for clients and improved quality of information provided to funding sources (Granello, Granello, & Lee, in press). Measuring treatment effectiveness need not be a difficult or cumbersome task. Simple measures of effectiveness can be implemented quite easily, and the demonstrated outcomes from such research can be a very effective tool for providing evidence of treatment success.

Methodological Considerations

To engage in outcome research, counselors must first have an understanding of the two main types of research that are used to demonstrate clinical success: efficacy studies and effectiveness studies. *Efficacy studies* use random assignment to treatment and control group, manualize treatment, and use participants who meet criteria for a single diagnosed disorder (Seligman, 1995; Wampold, 1997). Additionally, there are clearly defined inclusion and exclusion criteria for clients and an adequate sample size to obtain the necessary statistical power (Fishman, 2000). Efficacy studies provide useful information and are appropriate designs for laboratory studies or settings in which highly controlled manipulation of variables is possible (Sandell et al., 1997). However, these studies are very expensive and time-consuming and often are funded through a university or through a grant offered by a foundation or a pharmaceutical company.

Effectiveness studies, on the other hand, attempt to answer how well clients fare under treatment as it is actually practiced in the field. Such studies yield useful and credible information that can empirically validate psychotherapy (Lambert, Huefner, & Nace, 1997). Effectiveness studies recognize that less-than-methodologically-ideal situations exist in the field. Among these situations are that (a) therapy is not always of fixed duration, and typically continues until the client improves or quits or until insurance coverage runs out; (b) psychotherapy often is eclectic rather than manualized and typically is self-correcting (e.g., if one technique is not working, then another usually is tried); (c) clients typically present with multiple problems, some subclinical and some diagnosable, rather than the pure diagnoses represented in efficacy studies; and (d) psychotherapy in the field typically is concerned with improvements in general functioning rather than in specific symptom relief, which is the typical measure in efficacy studies (Seligman, 1995).

Efficacy and effectiveness studies have different strengths and limitations. Efficacy research typically has high internal validity but low external validity. The conditions under which efficacy research is conducted are so structured that there is a high degree of confidence that changes that occur are due to the treatment, not to confounding variables. However, the conditions under which efficacy research is conducted are often so dissimilar to what happens in the field that there is a low degree of confidence in generalizing the results of a particular study to field conditions. Conversely, effectiveness studies have high external validity but low internal validity. Because they sample a population directly from the field, there is a high level of confidence that results can be generalized to other members of the population (Fishman, 2000). The lack of a control group and of therapist adherence to specific treatment interventions are noteworthy, however, and lead to concerns about confounding variables (e.g., the passage of time) that might affect treatment results (Granello et al., 1999). Overall, efficacy and effectiveness studies provide complementary research designs. Counselors can use published efficacy studies to make initial choices about treatment interventions, then conduct effectiveness studies on their own practice to measure the success of their treatment (Granello & Hill, 2000).

Research Design

Research design is guided by the research questions under investigation (Granello & Hill, 2000). What specific information does the counselor wish to have about his or her practice or clients? Clinicians wishing to engage in tracking the success of a single client for reimbursement purposes would ask different research questions than would those wishing to investigate their treatment success with their overall client load or with clients having particular disorders (e.g., anxiety disorders).

Many effectiveness studies follow a pre-post or pre–post–follow-up design. That is, clients are given an instrument or series of instruments upon entering treatment, and the same instrument or instrument battery is given at discharge, and if desired, at pre-designated follow-up periods (typically 3, 6, or 12 months, or all three). Other types of effectiveness studies track the progress of a single client at various points in treatment (e.g., every week, every month), on a specific rating scale, with results that can be represented graphically to demonstrate progress. Still other studies use existing data from client records (e.g., Global Assessment of Functioning scores) to make comparisons over time or across client groups. Thus, for a single client, the counselor may choose to measure the reduction of a very specific symptom and engage in a single-case pre-post design, using a repeated measures t-test, or may choose to forego statistical analysis in favor of a graphic representation of multiple data points. To measure symptom reduction in multiple clients, the clinician may wish to collect demographic data and make comparisons (via repeated measures MANOVA) of reduction of various types of symptoms depending on demographic data (e.g., age, gender) or Axis I diagnosis. From this information, for example, a clinician could learn that he or she is very effective at helping clients with clinical depression to reduce their cognitive symptoms of depression but not as effective at helping to reduce the behavioral components. Likewise, she or he could discover that the treatments implemented seem to work well for female clients but are less successful with male clients. Clearly, all of this information can yield valuable data for improving clinical effectiveness.

Selecting Instruments

Instrumentation determines the type of data that can be obtained, and thus the choices regarding instrumentation must be made with care. The basic research questions that are being investigated should guide the instrument selection. Clinicians are strongly encouraged to use existing instruments with established validity and reliability whenever possible, rather than attempting to develop their own. Independently developed instruments require large commitments of time and resources to ensure reliability and validity, and once data is collected, no comparisons can be made with norming groups from existing research (Hansen, 1999). The test manual for a published instrument should provide norming samples that can help determine whether the person or sample being tested should be compared with the test norms. When selecting from existing instruments, practitioners should consider the cost of the instruments, including time required to administer, score, and analyze the results. Further, it is important to consider a measure that is sensitive to changes in symptomatology (Burlingame et al., 1995; Waxman, 1994; see Lambert, Ogles, & Masters, 1992 for methods to select and analyze the appropriateness of outcome instruments).

Using a small battery of instruments, rather than just one, may provide the best information. It may be useful to collect data from several different sources (e.g., client report, clinician rating, family/ teacher rating) to gain a clearer picture of the client's functioning (Sexton, Whiston, Bleuer, & Walz, 1997). Counselors should take care not to overburden their clients or to administer so many instruments that they are overwhelmed with data, however. Two or three short instruments, plus a demographics questionnaire, may be sufficient (Granello et al., 1999). Clinician ratings (e.g., a Global Assessment of Functioning score) can be an important component of treatment evaluation, as clinicians may be in a unique position to provide insight into patient progress. Using clinician ratings as a stand-alone measure of progress is unwise, however, as they have been criticized for their subjectivity (McLeod, 1994).

Using the Results

The results of effectiveness studies can be useful in a variety of ways. In several large-scale outcome studies conducted by the authors, data on program effectiveness were useful in marketing both adult and child partial hospitalization programs to the community, to insurance companies, and to managed care panels (Granello et al., 1999; Granello et al., in press). Importantly, a measure of client satisfaction was an essential part of this research and was highlighted in marketing materials. In a study of an eating disorder unit, results of the effectiveness research were used to increase hospital resources allocated to that unit (Granello & Hill, 2000).

Conducting such research has other, less tangible results. Clinicians with access to data can use those data to improve their treatment interventions, and research has found that practitioners' efficacy improves when they are involved in research (Hauri, Sanborn, Corson, & Violette, 1988). Reports from agencies that make systematic attempts to investigate their outcomes indicate that once clinicians become aware of variations in client outcomes, they are in a better position to generate ideas for improvement and hypotheses for further testing ("Authors pose," 1997). Thus, data collection and analysis may have great clinical importance.

Tips for Implementation

Although effectiveness studies clearly have limitations, we agree with Seligman's (1995) assertion that they are a complementary research method to efficacy studies. They provide practitioners with research that is clinically useful and important for negotiating managed care contracts, while allowing meaningful research to be conducted with minimal disruption to their work with clients.

Practitioners wishing to conduct outcome research in their own practice are encouraged to keep a few important suggestions in mind (see Granello et al., 1999 for a more complete discussion on implementation of effectiveness studies).

1. Effectiveness studies cannot be all things to all people. Complex designs with multiple administrations and a large number of instruments may so overwhelm the clinician that they are never completed or, once completed, are never statistically analyzed in a meaningful way. For practitioners just beginning to collect

data, our recommendation is to keep the data collection and analysis manageable.

2. Although outcome research need not be cost prohibitive, some foresight will be necessary to set aside sufficient funds for instruments and, if necessary, data analysis. We have found that university-agency collaboration, although not necessary, can provide a symbiotic relationship (data for the university, data analysis for the agency).
3. As much as possible, the collection of data should be integrated into clinical practice (e.g., put pretests in admissions packets so they are not forgotten).
4. For clinicians not currently collecting data, any step, however small, is a step in the right direction. Collecting data on treatment effectiveness can provide both an external benefit in terms of marketing and an internal benefit in validating and improving clinical success.

References

Authors pose 7 questions to address in designing outcomes system. (1997, August). *Behavioral Health Outcomes, 2*(7), 9–10.

Burlingame, G. M., Lambert, M. J., & Reisinger, C. W. (1995). Pragmatics of tracking mental health outcomes in a managed care setting. *Journal of Mental Health Administration, 22*, 226–236.

Cohen, L. H., Sargent, M. M., & Sechrest, L. B. (1986). Use of psychotherapy research by professional psychologists. *American Psychologist, 41*, 198–206.

Falvey, E. (1989). Passion and professionalism: Critical rapprochement for mental health research. *Journal of Mental Health Counseling, 11*, 86–95.

Fishman, D. B. (2000, May 3). Transcending the efficacy versus effectiveness research debate: Proposal for a new, electronic "journal of pragmatic case studies." *Prevention and Treatment, 3*, Article 8. [On-line]. Available: http://journals.apa.org/prevention/volume3/pre0030008a.html.

Granello, D. H., Granello, P. F., & Lee, F. (1999). Measuring treatment outcomes and client satisfaction in a partial hospitalization program. *The Journal of Behavioral Health Services and Research, 26*, 50–63.

Granello, D. H., Granello, P. F., & Lee, F. (in press). Measuring treatment outcome in a child and adolescent partial hospitalization program. *Administration and Policy in Mental Health.*

Granello, D. H., & Hill, L. (2000). Measuring treatment outcome in eating disorders programs: Recommendations for research design and implementation. Manuscript submitted for publication.

Hansen, J. C. (1999). Test psychometrics. In J. W. Lichtenberg & R. K. Goodyear (Eds.), *Scientist-practitioner perspectives on test interpretation* (pp. 15–30). Boston: Allyn & Bacon.

Hauri, P., Sanborn, C., Corson, J., & Violette, J. (1988). *Handbook for beginning mental health researchers.* New York: Haworth.

Havens, L. (1994). Some suggestions for making research more applicable to clinical practice. In. P. F. Tally, H. H. Strupp, & S. F. Butler (Eds.), *Psychotherapy research and practice: Bridging the gap* (pp. 88–98). New York: Basic Books.

Lambert, M. J., Huefner, J. C., & Nace, D. K. (1997). The promise and problems of psychotherapy research in a managed care setting. *Psychotherapy Research, 7,* 321–332.

Lambert, M. J., Ogles, B. M., & Masters, K. S. (1992). Choosing outcome assessment devices: An organizational and conceptual scheme. *Journal of Counseling and Development, 70,* 527–532.

McLeod, J. (1994). *Doing counselling research.* London: Sage.

Mirin, S. M., & Namerow, M. J. (1991). Why study treatment outcome? *Hospital and Community Psychiatry, 42,* 1007–1013.

Norcross, J. C. (2000, August 25). Toward the delineation of empirically based principles in psychotherapy: Commentary on Beutler (2000). *Prevention and Treatment, 3,* Article 28. [On-line] Available: http://journals.apa.org/prevention/volume3/pre0030028c.html.

Norcross, J. C., & Prochaska, J. O. (1983). Clinicians' theoretical orientations: Selection, utilization, and efficacy. *Professional Psychology, 14,* 197–208.

Plante, T. G., Anderson, E. N., & Boccaccini, M. T. (1999). Empirically supported treatments and related contemporary changes in psychotherapy practice: What do clinical ABPPs think? *Clinical Psychologist, 52,* 23–31.

Plante, T. G., Couchman, C. E., & Diaz, A. R. (1995). Measuring treatment outcome and client satisfaction among children and families. *Journal of Mental Health Administration, 22*, 261–269.

Sandell, R., Blomberg, J., & Lazar, A. (1997). When reality doesn't fit the blueprint: Doing research on psychoanalysis and long-term psychotherapy in a public health service program. *Psychotherapy Research, 7*, 333–344.

Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The *Consumer Reports* study. *American Psychologist, 50*, 965–974.

Sexton, T. L. (1996). The relevance of counseling outcome research: Current trends and practical implications. *Journal of Counseling and Development, 24*, 590–600.

Sexton, T. L. (2000). Restructuring clinical training: In pursuit of evidence-based clinical training. *Counselor Education and Supervision, 39*, 218–227.

Sexton, T. L., Whiston, S. C., Bleuer, J. C., & Walz, G. R. (1997). *Integrating outcome research into counseling practice and training*. Alexandria, VA: American Counseling Association.

Sherman, C. F. (1992). Changing practice models in managed health care. *Psychotherapy in Private Practice, 11*, 29–32.

Wampold, B. E. (1997). Methodological problems in identifying efficacious studies. *Psychotherapy Research, 7*, 21–43.

Waxman, H. M. (1994). An inexpensive hospital-based program for outcome evaluation. *Hospital and Community Psychiatry, 45*, 160–162.

Whiston, S. C., & Coker, J. K. (2000). Reconstructing clinical training: Implications from research. *Counselor Education and Supervision, 39*, 228–253.

About the Authors

Darcy Haag Granello, is an associate professor of counselor education at The Ohio State University. She received her Ph.D. in counselor education from The Ohio University and her M.S. in mental health counseling from Stetson University in DeLand, Florida. She has two main research interests: conducting outcome research in clinical mental health and assessing and enhancing the cognitive development of counselor trainees. She has published more than 30 refereed journal articles in national journals and five book chapters, including nine articles or book chapters related to conducting outcome research. She received a research award from ACES in 1998, and is the list owner for COUNSGRADS, the national student mailing list server.

Paul F. Granello, is an assistant professor of counselor education at The Ohio State University. He received his Ph.D. in counselor education from The Ohio University in Athens, Ohio, and his M.S. in mental health counseling from Stetson University in DeLand, Florida. His two primary areas of research are integrating wellness into mental health counseling and conducting outcome research. He has published 12 articles in national journals and five book chapters, including eight articles or chapters related to outcome research. He is a member of The Ohio State University College of Education Research Committee.

Chapter Fifteen

Communicating Assessment Results in the Counseling Interview

Albert B. Hood

Abstract

Clients who receive test interpretations generally make greater gains than those who do not. Yet clients' recall and understanding of test interpretations are frequently incomplete or inaccurate. A client-centered approach to testing and interpretation is described. The client first participates in the selection of the general type of tests to be administered, then the client plays the role of test interpreter, with counselor guidance. This process reduces the chance of a client misunderstanding test results or recalling them inaccurately.

The clinical use of psychological tests is typically included as one of the requirements in the graduate counseling curriculum through which the counselor-in-training is expected to become at least minimally proficient in the areas of test selection, evaluation, administration, and interpretation. There are test manuals, much testing literature, and good textbooks dealing with these subjects. An equally important subject—the communication of assessment results—typically receives scant attention, even though counselors are constantly required to interpret assessment results both to clients and to others such as parents, agencies, and other professionals. Effective communication is especially critical for counselors because in the end it is the understanding by the client or other individual who will be making decisions based on the results that will determine the actual application, if any, to which the assessment results will be put.

Most of us have acquaintances who have told us that their guidance counselor recommended on the basis of aptitude tests that they take a vocational program in carpentry or another of the skilled trades, but instead they became a social worker, a physician, or a college professor and are very satisfied and highly successful in this totally different career. In fact, they were probably not specifically told when they took the tests back in high school that they should become a carpenter, and their recall is probably quite selective. That it is not uncommon for former clients not only to mistake the type of assessment (ability, interest, or personality), but also to remember results selectively or erroneously, or to interpolate the word *should* into their interpretation—only emphasizes the importance of adequate client understanding (Zytowski, 1997).

Test Interpretation Research

How test results are reported or interpreted to clients and the accuracy of client understanding are extremely important but seldom studied aspects of counseling. Goodyear (1990) provided one of the first critical reviews of the literature in this area. He reported that studies generally show that clients who receive test interpretations—regardless of format and the particular outcome criteria employed—do experience greater gains than do those in control conditions. An interpretation of test results to clients, then, generally has a positive effect. A study by Finn and Tonsager (1992) provided support for Goodyear's conclusion. When they compared attention-control participants with clients who received Minnesota Multiphasic Personality Inventory (MMPI-2: Butcher et al., 1989) interpretations, those receiving the interpretations showed increased self-esteem and optimism about being able to overcome their problems as well as a decrease in symptoms. Goodyear found little research evidence that outcomes are differentially affected by the type of interpretation employed but reported that studies have shown that clients generally prefer individual integrative interpretations over self-interpretations or group interpretations.

Most of the studies identified by Goodyear (1990) measured outcomes over relatively short periods, such as two weeks after test interpretation. Furthermore, most of the research studies on the interpretation of test results have been limited to career counseling and have been conducted with either high school or college students as subjects. Virtually none have been concerned with personal counseling, psychotherapy, or with couples and family counseling, even though test interpretations are often employed with such clients. The outcome criteria used in many of the studies dealt with the memory or recall of test results and were based on the major assumption that increased self-knowledge was not only desirable but also helpful for the client.

In general, studies of the accuracy of recall of interest inventory results have not been encouraging (Froehlich & Moser, 1954; Zytowski, 1997). Correct recall of interest inventory results has ranged from 13% to 98%. Only 36% of college students contacted a year after receiving their Strong Interest Inventory (SII) results correctly remembered their highest general occupational theme score, and only 56% remembered their highest basic interest score.

In one study, students who had SII profiles interpreted to them in group settings were followed up a year later by means of a telephone interview (Hansen, Kozberg & Goranson, 1994). Respondents were asked about their recollections of their Holland code types, the high scores on the basic interest scales, and the occupational scales on which they received some of their highest scores. Of these students, 38% recalled the Holland theme with the highest interpretative comment, 62% recalled one of their six highest basic interest scores, and 77% recalled an occupational scale score that was in the moderately similar or higher range. Across all survey questions, the average recall accuracy was approximately 50%. More intelligent clients remembered their test results with greater accuracy—no surprise, as it would be expected that brighter individuals would have better recall of any information.

Early research did not seem to yield results that endorsed one type of test interpretation format over another (Forster, 1969; Gustad & Tuma, 1957; Rogers, 1954). Client preferences, however, generally favor interpretations conducted individually by the counselor, and these have been found to be the most effective in terms of favorability of client outcome (Oliver & Spokane, 1988). More recent studies clearly indicate that clients prefer integrative individual counseling as more attractive than test-centered individual or group interpretations (Miller & Cochran, 1979; Oliver, 1977; Rubinstein, 1978). The former format, of course, is considerably more expensive.

Studies have shown that clients generally accept positively worded interpretations more readily than negatively phrased ones (the Pollyanna effect; Sundberg, 1955). In addition, when the interpretation is value laden—for example, when abilities scores are the focus of the interpretation—that score is seen more positively and is more likely to be remembered (Dickson & Kelly, 1985). A problem with such studies is that they make no distinction between remembering and accepting test results. Most studies of test interpretation have used client recall of test results as the outcome criterion. The actual understanding or use of recalled results has seldom been investigated, although Goodyear (1990) did find several studies examining the accuracy of self-estimates of the characteristics measured by a test before and after test interpretation.

Several studies have dealt with so-called Barnum interpretations—generalized interpretations that often receive much credibility, such as those often found in horoscopes and astrological "personality profiles"

(Dickson & Kelly, 1985; Merrens & Richards, 1970). Barnum statements fall into several categories: the double-headed statement, the modal statement descriptive of virtually anyone, the vague statement, and the favorable statement. Studies of Barnum-effect statements have focused exclusively on personality interpretations and have found such statements enjoy exceedingly high client acceptance rates. In fact, clients often perceive generalized or even false test feedback based on astrology or Barnum-type interpretations to be more accurate than bona fide results. Several personality variables seem to be related to the acceptance of generalized personality interpretations, but there are no gender differences in this acceptance.

Accuracy of recall may be increased by providing more opportunities for depth of processing by the client. If the client actively forms many semantic associations with both new and old information during the interpretation process, this deeper level of processing should result in greater memory for the information. Clients are encouraged to actively connect the results to their own existing self-knowledge and potential career plans. The use of additional materials such as career resource books or other assessment tools also encourages the formation of such associations.

Principles of Test Interpretation

In counseling it is important to remember that there is almost always an implicit future orientation, even though the immediate goal is to help clients to make a particular decision or to understand themselves better. There is a belief that it is important for people to know themselves better because ultimately the self-knowledge gained in counseling and testing will enable them to have more effective and satisfying lives and to make wiser and more realistic plans.

It is necessary to have a thorough understanding of tests, particularly their theoretical foundations, if a counselor is to function as a professional in the test interpretation process rather than as a technician using a simple cookbook approach. Because tests are used to diagnose and predict, interpretations on the part of both the counselor and client must lead to the desired understanding and results. It must be remembered that a huge number of factors are involved in producing a particular test score. These include the clients' inherited abilities; their educational, cultural, family, and other experiences; their experiences with other tests, particularly psychological tests; their motivation; their test anxiety; the physical and psychological conditions under which they took the test; and the random variation in the test itself.

Various types of validity become extremely important in the interpretation of tests. It is therefore important to understand the construction and development of the test as well as its validity as determined by its relationship to that aspect of the construct which it is purported to measure. In every kind of test interpretation, it is assumed that there is a definite relationship between the person's score or result on a test and what it is being related to. Often this relationship is expressed in statistical terms such as correlation coefficients, descriptive and comparative statistics, or expectancy tables. To ensure client understanding, these concepts need to be explained in clear, understandable language.

Assessment results may be communicated through a variety of modes—with counselor interaction individually; in a small group with discussion; or in a large group with little or no discussion and without counselor interaction, as in the case of a score report or profile with a printed explanation, a narrative report, or a video or computer interactive supplement (Goodyear, 1990). An interactive approach is to be preferred, as shown by a study that compared a counselor-delivered interpretation with a counselor-client interactive interpretation (Hanson, Claiborn & Kerr, 1997). Clients not only preferred the interactive approach but also perceived the counselors to be more influential, expert, and trustworthy.

Client Participation

One finding that has stood out in studies of test interpretation is the value of client participation (Dressel & Mattson, 1950; Goodyear, 1990). Client participation in the selection of tests emphasizes that testing is an integral part of the counseling process and not an interruption of it. Most people generally approach tests—particularly aptitude and achievement tests—with some anxiety caused by fear of failure. Even interest and personality tests can reveal aspects of a person's attitudes and personality that indicate weaknesses or undesired alternatives and therefore may also be seen as something of a threat. Anxiety regarding testing is likely to carry over to the entire counseling process and certainly to receiving the results in an interpretation interview. If clients assist in the selection of the tests, they are more likely to be convinced of their usefulness and therefore be more motivated to do their best on ability tests and to be accurate and truthful in responding to items on interest and personality inventories. Having participated in the decision to use the tests, clients can be more objective in their perception of the results of the tests. They are also more likely to accept the results and their interpretations with less defensiveness (Fischer, 1970).

In the case of vocational counseling, clients tend to be dependent and test oriented. This emphasis on the test is a problem that often confronts educational and vocational counselors. Participation in test selection may also discourage the client from becoming dependent on the counselor, because the client accepts some responsibility for the selection of the testing instruments. A client's reactions to the suggestions and descriptions of the various tests can also provide useful information to the counselor. If the counselor is sensitive to the client's feelings about various aspects of the testing, these perceptions can lead to an informative examination during the interview or in a later session. Client participation in test selection may also lead to the selection of more appropriate tests because clients can help counselors understand what they already know and what they need to know.

There are, of course, situations in which tests are administered as part of a testing program and in situations apart from the counseling process, but the results are then used in counseling. In such cases it is obviously impossible to include clients in the selection of the tests. Here it becomes important for the counselor to communicate to clients and to determine whether the clients' interpretations indicate their understanding and insight.

In the test selection process the counselor needs to communicate the general role of the tests, the general procedures used, and the particular information being sought. Testing goes along with the total counseling process during this interview and should not be its only focus. Generally the client does not decide which specific test is the best measure, because this is a technical decision that counselors make on the basis of their professional knowledge. Instead, the counselor and client agree on the types of tests that will be the most useful and will provide information that is valid for whatever actions or decisions are going to be made. In general, clients are not nearly as interested in the specific characteristics of the test as they are in the implications the results will have for them. Therefore, counselors should describe the types of tests in general terms, rather than overwhelming clients with lengthy, technical descriptions of the tests and the many aspects of psychological measurement that are related to the field.

In general, a client's initial perceptions about the need for testing should not necessarily be taken at face value. For example, a request for a personality test should result in an effort to explore the meaning of the request rather than simple acceptance of it. Rather than simply being curious about the results of a personality test, the client may be having some significant problems that he or she is reluctant to reveal, such as anxiety or depression, and may be indirectly asking for help, using the request for testing as an avenue to get at the major problem.

Another important principle to be used in test selection is that other sources of data should also be explored. Counselors can first attempt to explore with clients previous experiences that may provide relevant information and self-descriptions regarding what they know about themselves. Their recall of previous experiences provides a great deal of information that may rule out the need for particular tests and can add much to what the tests that are administered reveal.

Client Interpretation of Tests

Several approaches to test interpretation emphasize or rely completely on counselor interpretation. In the interpretive technique, the counselor presents the test data in objective terms and interprets the data for the client. For example:

> We have found the best indication of success in most college courses is how well you do in high school and how you rate on an academic ability test. You were in the upper 10 percent of your high school class and exceeded seven or eight out of ten college students on the academic ability test. Most people with scores like that learn complex things relatively easily and quickly. For example, 60 to 80 out of 100 students with scores like that get average grades or better in the three colleges to which you are considering applying.

Then there is the explanatory approach, in which the counselor interprets the results in a subjective, non-statistical personalized prediction:

> As far as I can tell from this evidence of aptitude, your chances of getting into medical school are poor, but your possibilities in business seem to be much more promising. Here are some of the reasons for my conclusions: you have done very poor work in zoology and chemistry. Your patterns of interests on the interest inventory are not characteristic of successful physicians, suggesting that your interests are unlike those of most of the folks in that field. On the other hand, you do well in mathematics, have good general ability, and your interests are like those of people in several business fields. These facts seem to me to argue for your selection of several options within the business field to explore.

Whether the counselor tends to be objective or subjective, the client is still in the position of receiving the counselor's interpretation of the results. The client interpretation approach, however, requires that the client play the role of test interpreter. This method can be employed with high school and college students and adults who come to counseling centers but is not meant for use with clients who are emotionally disturbed. The counselor's role in this approach is to prepare the client for the interpretation task and to give guidance and support during the process. The client's role is to learn the information necessary for interpreting the test results, to make the interpretation, to explore the ramifications of this interpretation, and to follow through by making decisions, adjusting plans, or otherwise implementing a modified self-perception.

Under this interpretation style, the counselor begins by defining the construct being evaluated; for example, vocational interest or mathematical aptitude or a personality characteristic. This general definition naturally leads to a discussion of what the test at hand is measuring. It is at this point that the counselor ensures that no fears or stereotypes are going to distort later perception of the test results. The discussion then moves to the subject of error of measurement and perhaps a brief explanation of what the test is *not* measuring. The counselor's judicious provision of information saves time and ensures that the client is receiving accurate information, that it is presented in a conversational style, and that there are frequent summaries. The counselor should also explore any comments the client makes. In this technique the counselor is counseling all the time and can shift out of the test interpretation role at any time and return to it later.

In the next phase, the counselor presents information regarding the manner in which the test results or profiles are presented. If relevant, there is discussion of the implications involved in making forced-choice comparisons. Percentile ranks are illustrated and norm groups carefully explained. The approach in this phase is active and Socratic. Again the counselor uses frequent summaries and single-question "quizzes" to ensure correct learning and verbally reinforces accurate client insights. The results are usually presented with a suggestion that the client study them for a while, then tell the counselor what they mean. At this point, the counselor's role shifts to clarification and exploration of the interpretations. This procedure ensures that the interpretations, evaluations, or biases are the client's. In this phase, by simple reflection or restatement, the counselor can help the client to clarify and elaborate on what the results mean for him or her. If the client appears to be faced with the problem of incorrect learning, then the counselor must repeat the introductory material accurately, as information that is incorrectly received is worse than no information at all. If the client interprets the results accurately but appears to be unhappy about the findings, the counselor can begin immediately to help the client work through the unhappiness.

The advantage of this approach is that it reinforces the client as a person who is capable of interpreting and understanding psychometric results. The chances of the client misinterpreting what the test actually measured are greatly reduced if not eliminated. The interpretations are the client's insights. As a result, the client is more adequately prepared to understand and assimilate the results following this type of interpretation. The client has assumed the main responsibility for interpretation, decision making, and planning for implementation of decisions made.

Interpreting test results over more than one session is desirable, in line with the notion that distribution of practice improves the acquisition of knowledge and the memory of information. In the first session of interpreting Strong Interest Inventory, the counselor might ask the client to predict on which of the occupational themes he or she is likely to have scored the highest and lowest. Then the actual results would be discussed. In the next session the client would be asked to recount what he or she remembered from the previous session and integrate this information with the basic interest scale results. In the subsequent session the counselor would encourage the client to recount what was discussed in the previous two sessions then would introduce the occupational scales. By thus reconciling discrepancies between preexisting beliefs and actual scores, the counselor could guide the client to increased acceptance of the Strong profile and thus greater memory for the results (Hansen et al., 1994).

Counselors in some settings must work within more limited time parameters, for example, high school counselors usually do not have the opportunity to spend four or five sessions interpreting one instrument. They may, however, have contact with a student over a number of years in different contexts, such as course selection, career exploration, or college selection. The depth of processing and distribution of practice approaches can be addressed through methods unique to the school counseling situation. Parents could also be involved in the interest exploration process. The testing results could also be integrated into a junior- or senior-level class on career exploration.

Conclusions

Psychological tests are used by personnel managers to hire employees, by school psychologists to track pupils, by clinical psychologists to diagnose patients, by college admissions staff to admit students, and by forensic psychologists to determine sanity. In the counseling setting, however, psychological tests are used to help clients understand themselves. Counselors use tests primarily to assist individuals in developing their potential to the fullest and to their own satisfaction. In this setting test results are designed to be used by the clients themselves, rather than by others making decisions on the clients' behalf. Thus, how adequately the clients themselves understand the test results is more important than the counselor's knowledge or understanding. With the prevalence of negative attitudes toward psychological tests, counselors may be reluctant to make adequate use of them in assisting clients, but they should remember that the use of tests in counseling differs from their use in other contexts. Test results in counseling constitute interventions that can facilitate change and can lead to greater awareness, knowledge, and self-understanding, which can result in clients' making better and more effective decisions.

References

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., Hathaway, S. R., & McKinlay, J. C. (1989). *Minnesota Multiphasic Personality Inventory–2*. Minneapolis, MN: University of Minnesota Press.

Dickson, D. H., & Kelly, I. W. (1985). The "Barnum effect" in personality assessment: A review of the literature. *Psychological Reports, 57*, 367–382.

Dressel, P. L., & Mattson, R. W. (1950). The effect of client participation in test interpretation. *Educational and Psychological Measurement, 10*, 693–706.

Finn, S. E., & Tonsager, M. E. (1992). Therapeutic effects of providing MMPI-2 test feedback to college students awaiting therapy. *Psychological Assessment, 4*, 278–287.

Fischer, C. T. (1970). The tester as co-evaluator. *Journal of Counseling Psychology, 17*, 70–76.

Forster, J. R. (1969). Comparing feedback methods after testing. *Journal of Counseling Psychology, 16*, 222–226.

Froehlich, C. P., & Moser, W. E. (1954). Do counselees remember test scores? *Journal of Counseling Psychology, 1*, 149–152.

Goodyear, R. K. (1990). Research on the effects of test interpretation: A review. *The Counseling Psychologist, 18*, 240–257.

Gustad, J. W., & Tuma, A. H. (1957). The effects of different methods of test introduction and interpretation on client learning in counseling. *Journal of Counseling Psychology, 4*, 313–317.

Hansen, J. C., Kozberg, J. G., & Goranson, D. (1994). Accuracy of student recall of Strong Interest Inventory results one year after interpretation. *Measurement and Evaluation in Counseling and Guidance, 26*, 235–242.

Hansen, J. C., Neuman, J. L., Haverkamp, B. E., & Lubinsky, B. R. (1997). Comparison of user reaction to two methods of Strong Interest Inventory administration and report feedback. *Measurement and Evaluation in Counseling and Development, 30,* 115–127.

Hanson, W. E., Claiborn, C. D., & Kerr, B. (1997). Differential effects of two test-interpretation styles in counseling: A field study. *Journal of Counseling Psychology, 44*, 400–405.

Merrens, M. R., & Richards, W. S. (1970). Acceptance of generalized vs. "bona fide" personality interpretations. *Psychological Reports, 27,* 690–694.

Miller, M. J., & Cochran, J. R. (1979). Evaluating the use of technology in reporting SCII results to students. *Measurement and Evaluation in Guidance, 12,* 166–173.

Oliver, L. W. (1977). Evaluating career counseling outcomes for three modes of test interpretation. *Measurement and Evaluation in Guidance, 10,* 153–161.

Oliver, L. W., & Spokane, A. R. (1988). Career intervention outcome: What contributes to client gain? *Journal of Counseling Psychology, 35,* 447–462.

Rogers, L. B. (1954). A comparison of two kinds of test interpretation interview. *Journal of Counseling Psychology, 1,* 224–231.

Rubinstein, M. R. (1978). Integrative interpretation of vocational interest inventory results. *Journal of Counseling Psychology, 25,* 306–309.

Sundberg, N. D. (1955). The acceptability of "fake" versus "bona fide" personality test interpretations. *Journal of Abnormal and Social Psychology, 50,* 145–147.

Zytowski, D. G. (1997). *How to talk to people about their interest inventory results.* Paper presented at the Society for Vocational Psychology conference, Bethlehem, PA.

About the Author

Albert B. Hood is professor emeritus at the University of Iowa. He obtained his Ed.D. at Cornell University. Among his research interests are psychosocial development of college students and psychological assessment. Hood is the author of more than 100 publications including five books, among them are *Key Resources on Student Services* (Jossey-Bass) and *Assessment in Counseling* (American Counseling Association). He also serves as editor of the *Journal of College Student Personnel*. A Fellow of the APA, Hood is recipient of the Contribution to Knowledge Award (ACPA) and the Extended Research Award (ACA). He also was awarded a Fulbright Lectureship in the Soviet Union.

Chapter Sixteen

Response to Surveys by Electronic and Traditional Means

David J. Lundberg

Abstract

In the summers of 1997 and 1998, a survey of AAC members was conducted via newsletter. Given a choice to respond via business reply mail or via various electronic means, 82% of members responded by mail, suggesting this may still be perceived as the easiest means of responding. Implications for the use of electronic surveys are discussed.

During the summer of 1997, a survey of Association for Assessment in Counseling (AAC) members was conducted, with a follow-up survey during the summer of 1998. A one-page (two-sided) survey was constructed to assess members' use of technology in counseling assessment. Survey questions on specific uses of known technology appropriate to the assessment process were developed following the standard definitions of assessment, measurement, evaluation, and interpretation (Payne, 1992; Vacc & Loesch, 1994). Questions on policies, procedures, and ethical standards in counseling assessment were also formulated. Because gathering information on effectiveness of technology and future needs was considered potentially valuable to association members and the counseling profession, questions were developed to assess these areas. The resultant survey form is located at www.ncat.edu/~schofed/aac. The results of the survey are available at jtc.colstate.edu/vol1_1/assessment.htm.

The survey was distributed via *Newsnotes* (the AAC quarterly newsletter) to the entire membership. Recipients were given the opportunity to respond in one of four ways: by business reply mail (a business reply envelope was included with the paper survey form in *Newsnotes*), by Internet website (an electronic survey form was available), by e-mail, or by fax. The two mailings went to an average of 2,130 readers, and 153 different individuals responded, resulting in an overall response rate of 7.2%.

Results and Discussion

Of the survey participants, 82% (126) responded by business reply mail. The remaining 18% (27) of the respondents replied via one of the electronic methods, with Internet website being the most popular electronic response (12% of total respondents). Approximately 4% of total respondents used e-mail, and 2% of survey participants used fax. The relatively small percentage of individuals using electronic methods may indicate that the familiar method of responding by pencil and paper was still regarded as the easiest way to participate.

Direct-mail surveys often have response rates of 30% or more (Heppner, Kivlighan, & Wampold, 1992) and have been a common and preferred method for survey research of geographically separated samples. Newsletter surveys typically have lower response rates, although information on this type of survey is less widely disseminated and discussed. Electronic surveys of various types are increasingly being used. Response rates, response bias, and other factors affecting results are less well known with electronic surveys than with the more traditional types of surveys.

Making electronic surveys easy to complete is probably just as important as it has always been with traditional surveys. Targeting the correct audience and using a personal appeal, as with direct mail and follow-up, is a time-tested and proven method of obtaining good response rates and representative samples. How to make electronic surveys personal and user-friendly is a crucial area for experimentation and learning. As we gain knowledge about the science and art of using electronic surveys effectively, response rates will surely increase. Perhaps then the efficiency of using technology for survey research will coincide with robust samples and good, generalizable results will follow.

Electronic surveys will not go away, but they will become better. After nearly a generation of using personal computers, a balance between "high touch" (interpersonal skills) and high technology (Harris-Bowlsbey, 1984), remains an essential ingredient in obtaining good results from human beings who communicate electronically.

References

Harris-Bowlsbey, J. (1984). High touch and high technology: The marriage that must succeed. *Counselor Education and Supervision, 24*, 6–16.

Heppner, P. P., Kivlighan, D. M., & Wampold, B. E. (1992). *Research design in counseling*. Pacific Grove, CA: Brooks/Cole.

Payne, D. A. (1992). *Measuring and evaluating educational outcomes*. New York: Macmillan.

Vacc, N. A., & Loesch, L. C. (1994). *A professional orientation to counseling* (2nd ed.). Muncie, IN: Accelerated Development.

About the Author

David J. Lundberg is assistant professor of counseling at North Carolina A&T State University in Greensboro, North Carolina. He holds a master's degree in counseling from Boston University and a Ph.D. in counselor education from the University of North Carolina–Greensboro. Primary research interests are in technology in counseling and multicultural counseling. He has published articles in the *Journal of Career Development*, the *Journal of Humanistic Education and Development*, and the *Journal of Technology in Counseling*. He currently serves on the Executive Council of the Association for Assessment in Counseling.

Chapter Eighteen

Using the Internet to Enhance Test Interpretation

James P. Sampson Jr.[1]

Abstract

The Internet offers both potential advantages and challenges in test interpretation. Several means for providing background information and test interpretation to test takers via password-protected websites, list servers, and videoconferencing are presented. The potential use of computer conferencing in supervision is also discussed. Potential ethical issues surrounding the use of the Internet to supplement or replace face-to-face interaction are reviewed.

Computer applications in assessment have been in use for more than forty years. Mainframe computers made it cost-effective to score, profile, and produce narrative interpretive reports for traditional paper-and-pencil tests. Personal computers subsequently made it cost-effective to add test administration and multimedia elements to these functions. The Internet is now adding the potential for remote delivery of test administration, scoring, profiling, report writing, and multimedia functions, as well as adding potentially cost-effective capability in communication and links to related information.

Principles of effective and responsible test use are embodied in testing standards (AERA, APA, & NCME, 1985; Joint Committee on Testing Practices, 1988; and AMECD, 1989) and assessment competency statements (Garfield & Prediger, 1994). These standards and competency statements refer to common elements of the assessment process that include test selection, administration, scoring and interpretation, and communicating effectively with test takers and parents or guardians in the case of minors. In practice, these testing elements can be sequenced as follows: (a) selection, (b) orientation, (c) administration, (d) scoring, and (e) interpretation. The focus of this article will be on using the Internet to enhance test interpretation. I begin with a review of potential Internet applications in test interpretation and conclude with issues associated with Internet use in test interpretation.

Internet Applications in Test Interpretation

One important potential advantage of using the Internet as a test interpretation resource is the ease with which the interpretive data can be kept current. The costs associated with disk manufacture, shipping, and billing make it expensive for publishers to update interpretive software as each new enhancement becomes available from ongoing practice and research. As a result, test publishers tend to wait until considerable knowledge or software enhancements accumulate before releasing a new version of the interpretive software. With the Internet, test publishers can update software as the information becomes available, notifying users of recent changes as they access a password-protected site and billing on a per-use or an annual license basis.

The Internet potentially can be used in a variety of ways to enhance test interpretation. This section will deal with client preparation for test interpretation, generalized test interpretation, specific test interpretation, and supervision. All of the following applications can be accomplished with technology that is currently available.

Client Preparation for Test Interpretation

Effective test interpretation actually begins before a test is administered. Orientation to testing provides a foundation for delivering a subsequent test interpretation. Problems in test interpretation often can be prevented if clients (and their parents or guardians, if appropriate) are adequately informed of the purpose and process of testing. For example, the common client misperception that the "scientific" nature of testing will provide an "answer" to his or her problem can potentially be corrected by information delivered during orientation. Intelligent counselors, however, frequently become bored with repetitive tasks and are attracted to more intellectually challenging and interesting tasks.

The problem is that test orientation often involves presenting repetitive information. Because computers do not become bored with repetitive information delivery, they are likely to be an effective resource for learning general principles of test orientation. Years of experience in delivering computer-assisted instruction could be easily applied to the task of test orientation on the Internet. Using a password to maintain security, clients could access the orientation at their convenience at home or at a public location such as a public library.

The counselor would add any necessary client-specific orientation that is not covered by computer-assisted instruction.

Generalized Test Interpretation

Like test orientation, test interpretation can often be repetitive and time-consuming for the counselor. As a result, counselor performance may be compromised. Two negative outcomes may result for the client. First, the client may not receive necessary information regarding basic terminology used in a particular test and what is being measured by scale and total scores. This lack of basic knowledge may make it more difficult for the client to understand and apply the specific interpretive information provided by the counselor. Second, if the counselor appears bored while delivering basic information, the client may misperceive that the counselor is bored with him or her, and the counseling relationship may be harmed as a result. Even if the counselor does a good job of communicating basic concepts, the time spent in this way means that less time is available to help a client gain insights about factors that influence his or her behavior and to help integrate insights gained in assessment into a realistic plan for behavior change. Using a computer to provide a generalized interpretation of test results can help a client to be better prepared for a specific test interpretation by "being aware of basic terminology, concepts, and the general nature of their scores" (Sampson, 1983, p. 294). By allocating the repetitive computational and instructional tasks to the computer, the counselor can focus on interpersonal functions associated with helping clients understand and apply test results to their individual circumstances (Sampson, in press, a).

Specific Test Interpretation

Building on the foundation of the generalized test interpretation, specific test interpretation adds interpretation of individual scales and aggregate score profiles as well as recommendations for action based on test results. In the case of self-assessment instruments, such as the Self-Directed Search (SDS, Holland, 1994), the measures are designed to be administered and interpreted without input from a counselor. As a result, self-assessment instruments may be delivered on the Internet by using or adapting existing personal computer-based interpretations, such as the interpretation for the Self-Directed Search (Reardon & PAR Staff, 1996). Although self-assessment instruments can be used without counselor input, Reardon and Lenz (in press) noted that experience with the SDS has shown that counselor input enhances the effectiveness of interpretations. In the case of counselor-mediated (traditional) assessment, the measures require that trained practitioners deliver test interpretations to clients. This section will deal with computer-based test interpretation, two-way videoconferencing, moderated list servers, computer-moderated conferences for group interpretation, and follow-up resource links.

Computer-Based Test Interpretation

Computer-Based Test Interpretation (CBTI) can enhance the validity and reliability of testing by providing counselors with an expanded and consistent knowledge base for test interpretation. Accumulated research data and practitioner experience expand the knowledge base for interpretation, while the standardized nature of computing contributes to the consistency of interpretation. In comparison with practitioner-developed reports, CBTI reports tend to be more comprehensive and objective and less subject to interpreter bias (Sampson, in press, a). Varying types of CBTI exist according to the type of knowledge base that is used for the software. CBTI may be categorized as descriptive, clinician-modeled (renowned clinician type), clinician-modeled (statistical model type), and clinical actuarial (Roid & Gorsuch, 1984). CBTI has also been categorized into three levels: (a) the statement level contains data-based descriptions; (b) the narrative level adds the judgment of experts in sequencing interpretive statements; and (c) the decision level adds prediction of client behavior (Lanyon, 1987).

Computer-based test interpretation via the Internet can be used in three different modes. When using self-assessment instruments, clients can independently access CBTI from password-protected Internet sites immediately after test administration is complete. Given that the Self-Directed Search was designed to be used with little or no counselor intervention (Reardon & Lenz, in press), the SDS could be administered and interpreted over the Internet. In this case, generalized and specific test interpretation are combined for the user. When using counselor-mediated assessment, the client first reviews a generalized test interpretation, then discusses his or her results with a counselor (face-to-face or via a videoconference over the Internet), and finally reviews a specific test interpretation delivered from a password-protected Internet site as a homework assignment. Some counselors might prefer for clients to review both the generalized and the specific interpretation as preparation for counseling. In the first case, the specific interpretation on the Internet reinforces learning that occurs in counseling, whereas in the second case, specific interpretation serves as an advance organizer for subsequent learning occurring in counseling. The current narrative interpretive report for the Strong Interest Inventory (SII, Hansen, Harmon, Borgen, & Hammer, 1994) could be delivered over the Internet in this manner. The third mode for delivering CBTI occurs when the principal consumer of test data and reports is the counselor, rather than the client. In this case, no generalized interpretation is provided to the client, and the counselor accesses a specific interpretation from a password-protected Internet site. For example, the current narrative interpretive report for the MMPI-2 (Butcher et al., 1989) could be delivered over the Internet in this manner.

By integrating CBTI, multimedia, and the Internet, it will be possible to better attend to multicultural issues in test interpretation. The gender, age, race, and ethnicity of the individual visually presenting information on test interpretation can be made to match the group membership of the test taker. Keeping group membership constant should make it easier for the client to relate to and understand the individual presenting the interpretation (Sampson, 1990). Additional multicultural research on test content and test interpretation can be added to CBTI as the research becomes available.

Much of the current Internet is text intensive, data intensive, and structured. These characteristics make it difficult for the many individuals with limited literacy skills to access and successfully use the Internet. Integrating CBTI and multimedia can make it easier for individuals with limited literacy skills to use the Internet. Providing versions of test interpretations with more video content and less text has the potential to help both individuals with limited literacy and individuals with a predominately visual learning style.

Two-Way Videoconferencing

Presentation and discussion of test interpretations could be accomplished via two-way videoconferencing over the Internet. This use of technology may be especially appropriate for clients in geographically remote locations and clients with physical disabilities who could choose to receive services at their residence. E-mail could be used to schedule test-interpretation sessions. Documentation of the completion of a test interpretation, including all test reports and intervention (treatment) plans, could be automatically added to the client's case notes. Subsequent client questions or concerns could be e-mailed to the counselor for immediate response or discussion at the next scheduled counseling session (Sampson, Kolodinsky, & Greeno, 1997).

Delivering CBTI via two-way videoconferencing could be an option that individuals who have completed self-assessment measures might select. CBTI for self-assessment could indicate the availability and potential benefits of two-way videoconferencing with a counselor trained in interpreting the specific test. Individuals selecting this option could either pay for the time used (as is the case for many current telephone help lines for computer software) or the service could be paid for by an organization with a mission to serve a specific population (such as employment service staff helping individuals make the transition from welfare to work). Self-assessment measures are increasingly being incorporated into comprehensive counseling resources, such as computer-assisted career-guidance systems. Two-way videoconferencing makes it possible for the counselor to intervene "in the moment," providing access to client perceptions and behavior as they occur, instead of subsequently discussing a client's reconstruction of a learning event that has occurred in the past (Sampson et al. 1997).

Moderated List Servers

Moderated mailing lists assist individuals with common interests to communicate with each other (Offer & Watts, 1997). A list server allows sequential public exchange of text-based information on a predetermined topic among a predetermined group of individuals (Sampson et al., 1997). Messages are distributed to all individuals who have registered to participate on the list server. The lack of interaction in real time is offset by the convenience of being able to view messages at any time. The participant may choose to post messages or simply to read the available messages and maintain a degree of anonymity. The moderator is responsible for keeping the interaction focused, halting inappropriate information exchanges, and proactively dealing with potential ethical problems. This resource would allow a counselor to respond to general questions about test interpretation or specific questions about individual test results (assuming that informed consent has been given when joining the list server). Participants potentially can learn from the interpretive insights revealed by the counselor and other members of the list server. The list server also can provide some social support for confronting issues revealed in test results and in taking action for positive behavior change.

Moderated Computer Conferences

A moderated computer conference allows simultaneous public interaction among a predetermined group of individuals (Sampson et al., 1997). The requirement of adhering to a scheduled time for interaction is offset by the higher level of interpersonal interaction that is possible in real time. The group dynamics associated with group counseling are operative but not identical in computer conferences. As a result, the availability of a moderator may help keep the interaction among participants focused, ensure that all participants have the opportunity to contribute, halt inappropriate information exchanges, and proactively deal with potential ethical problems. A moderated computer conference can serve the same functions as the moderated list server described previously with the exception that real time interaction is possible, providing opportunities to explore issues more quickly and in greater depth and to provide a higher level of social support.

Follow-up Resource Links

A potentially powerful feature of the Internet is the ability to use links embedded in one Internet website to access related information and services available at other Internet websites. Links can be used to promote additional learning related to test interpretation. For example, an interest inventory interpretation could provide links to occupational information websites for the occupations on which the client scores highly as a stimulus for career exploration. Similarly, the interpretation of a measure of study-skill behaviors could provide links to websites delivering specific study-skills instruction.

Supervision of Test Interpretation

The Internet has expanded opportunities for delivering supervision, potentially enhancing both the quantity and quality of interaction (Casey, Bloom, & Moan, 1994). The Internet can be used to facilitate supervision in several ways. A moderated list server could be used as a form of group supervision, with counselors requesting assistance for difficult interpretive issues. The moderator could be selected on the basis of specific interpretive expertise as well as his or her group facilitative skills. In this case, the role of the moderator would be expanded to include sharing his or her interpretive expertise and ensuring that the contributions of the other participants are appropriate for a specific test. Two-way videoconferencing is another possible means for individual supervision. The client's case notes (Casey et al., 1994), test results, and CBTI report could be attached to an e-mail file and sent to the supervisor to help with preparation for supervision. The supervisor and supervisee could then discuss a specific test interpretation in depth from remote locations in real time. A variation might include adding a consultant from a remote location to an ongoing supervisory relationship when an unusual interpretive question requires highly specialized expertise.

Issues Associated with Internet Use in Test Interpretation

Although the Internet applications described here offer the potential to enhance the access to and the quality of testing, issues also exist that have the potential to nullify the potential benefits of using this technology. This section deals with inappropriate use of counselor-mediated assessment, relationship issues, ethics, credentialing, and counselor training.

Inappropriate Use of Counselor-Mediated Assessment

As stated previously, counselor-mediated (traditional) assessment is designed to involve trained practitioners in delivering test interpretations to clients, whereas self-assessment measures are designed to be self-interpreted. Problems occur when counselor-mediated measures are delivered on the Internet without practitioner intervention. The first problem is the assumption that the results of the Internet-delivered version are equivalent in validity to the results of the traditional measure. The second problem is the assumption that the written interpretation offered on the Internet is equivalent in validity to the interpretation offered by a practitioner. Moving the interpretation of a self-assessment measure to the Internet is appropriate if the measure was originally validated to be self-interpreted. Unless validation data are available, the interpretation of counselor-mediated measures on the Internet without practitioner intervention is inappropriate (NCDA, 1997).

Relationship Issues

Videoconferencing and face-to-face interaction have been shown to be similar but not identical forms of communication. In comparison with face-to-face interactions, videoconferencing results in a more intense task focus and greater participant awareness of their physical appearance in the visual recording process (Oravec, 1996). The question, "Is remote videoconference interaction between a counselor and a client in a helping relationship equivalent to face-to-face interaction?" is an interesting but not crucial question. Given our current knowledge, the ultimate answer will likely be no, that videoconferencing and face-to-face interaction are different forms of communication. The more important question is, "Does remote videoconference interaction between a counselor and a client in a helping relationship assist clients in understanding and applying test results to solving problems and changing behavior?" Development of initial Internet applications and subsequent research on effectiveness is necessary to maximize the benefits and minimize the limitations associated with using this technology in counseling (Sampson, in press, b).

Ethical Issues

Numerous ethical issues have been raised relative to delivering assessment information and counseling over the Internet (Bartram, 1997; Sampson, in press, b; Sampson et al., 1997). The confidentiality of client data transmission and storage of assessment data may be compromised. It is also possible to deliver interpretive information via the Internet that is attractively presented but inherently invalid. There may be a lack of counselor intervention for clients who need a more personalized level of assistance. Inadequately trained or overworked counselors may misuse or become dependent on software such as CBTI. A lack of counselor awareness of important location-specific circumstances may cause a counselor in a remote location to misinterpret client data or fail to recognize relevant issues. Clients with limited financial resources may have difficulty gaining access to the Internet. Finally, accessing the Internet from a residence shared with other individuals may not provide the auditory and visual privacy necessary for the client to establish and maintain a counseling relationship.

Initial issues associated with computer networking were addressed in ethical standards and practice guidelines adopted by the American Association for Counseling and Development (AACD, 1988), the National Board for Certified Counselors (NBCC, 1989), the National Career Development Association (NCDA, 1991), and the American Psychological Association (APA, 1986). These initial standards on computer networking have recently been revised and expanded to deal specifically with the provision of information and counseling services over the Internet.

The National Board for Certified Counselors and Council for Credentialing and Education (NBCC & CCE 1997) webcounseling standards contain links to existing standards regarding confidentiality, supervision, relationship issues, release of information, record keeping, self-disclosure, certification and licensure, research, informed consent, impostor clients and counselors, security, local counseling support, liability, counselor access off-line, inappropriate presenting concerns, assessment and intake, communication problems, and relationship issues. The NCDA (1997) Internet standards specifically deal with the qualifications of the developer or provider, access to Internet sites, counselor understanding of local environment, content of career counseling and planning services, appropriateness of the client for receipt of services, appropriate local support for the client, clarity of the contract with the client, inclusion of linkages to other websites, use of assessment, job posting and searching, and unacceptable counselor behaviors.

Credentialing

The Internet poses some important challenges regarding credentialing. At present, it is uncertain how state counselor licensure laws will apply to a counselor delivering information and services out of state (Sampson et al., 1997). The same issue applies to delivering interpretive information and services across national boundaries (Bartram, 1997). Counselors delivering interpretive information and services over the Internet need to clearly indicate their credentials, including the complete name of the credential and the name and address of the credentialing organization. Existing Internet websites often fail to indicate the credentials of the service provider (Sampson et al., 1997). The potential lack of client awareness of the role of credentialing in protecting the public encourages unqualified persons to offer assessment information and services.

Counselor Training

Preservice and in-service counselor training is essential if counselors are to use the Internet effectively to serve their clients. Both students graduating from counselor-preparation programs and experienced counselors need to be competent in using Internet search engines, familiar with current websites related to counseling, skilled in evaluating the quality of websites, competent in integrating counseling interventions with Internet use, knowledgeable of the process for implementing Internet applications into counseling services, and aware of ethical issues and related professional standards. Students in training and practicing counselors who wish to take leadership in developing Internet applications need to supplement this preparation with instructional-design competencies and website-design skills.

Conclusion

Although the move to a paperless society has been less rapid than some futurists predicted, there appears to be inexorable movement in the direction of increased Internet use. Several factors are encouraging this trend. First, the cost-effectiveness of computer technology continues to improve dramatically. Second, Internet applications in general are growing exponentially despite the fact that the majority of Americans still do not have Internet access at home. Third, the pressure for distance learning will continue as the lifelong demand for education and training increases and funding remains limited. Appropriate distance-learning choices will increasingly be made on the basis of distance guidance, and testing will likely continue to play an important role in the guidance function. The speed at which these changes will occur can be debated, but the general direction of the change seems clear. It would seem wise to experiment carefully with appropriate applications of this technology and to deal proactively with potential limitations while there is still time to shape the early adoption of the Internet in testing.

References

AACD. (1988). Ethical standards. *Journal of Counseling and Development, 67,* 4–8.

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1985). *Standards for educational and psychological testing.* Washington, DC: APA.

APA. (1986). *Guidelines for computer-based tests and interpretations.* Washington, DC: Author.

Association for Measurement and Evaluation in Counseling and Development (AMECD). (1989). *Responsibilities of users of standardized tests.* Alexandria, VA: AACD.

Bartram, D. (1997). Distance assessment: Psychological assessment through the Internet. *Selection and Development Review, 13,* 15–19.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., Hathaway, S. R., & McKinlay, J. C. (1989). *Minnesota Multiphasic Personality Inventory–2.* Minneapolis, MN: University of Minnesota Press.

Casey, J. A., Bloom, J. W., & Moan, E. R. (1994). Use of technology in counselor supervision. In L. D. Borders (Ed.), *Counseling supervision.* Greensboro: University of North Carolina at Greensboro, ERIC Clearinghouse on Counseling and Student Services. (ERIC Document Reproduction Service No. ED 372 357)

Garfield, N. J., & Prediger, D. J. (1994). Assessment competencies and responsibilities: A checklist for counselors. In J. T. Kapes, M. M. Mastie, & E. A. Whitfield (Eds.), *A counselor's guide to career assessment instruments* (3rd. ed., pp. 41–47). Alexandria, VA: NCDA.

Hansen, J. I., Harmon, L., Borgen, F., & Hammer, A. (1994). *Strong Interest Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Holland, J. L. (1994). *Self-Directed Search*. Odessa, FL: Psychological Assessment Resources.

Joint Committee on Testing Practices. (1988). *Code of fair testing practices in education*. Washington, DC: APA.

Lanyon, R. I. (1987). The validity of computer-based personality assessment products: Recommendations for the future. *Computers in Human Behavior, 3*, 225–238.

NBCC. (1989). *Code of ethics*. Alexandria, VA: Author.

NBCC & CCE. (1997). *Standards for the ethical practice of webcounseling*. Greensboro, NC: Author.

NCDA. (1991). Ethical standards. In D. Engels (Ed.), *The professional practice of career counseling and consultation: A resource document* (2nd. ed., pp. 26–33). Alexandria, VA: Author.

NCDA. (1997). *NCDA guidelines for the use of the Internet for provision of career information and planning services*. Alexandria, VA: Author.

Offer, M., & Watts, A. G. (1997). *The Internet and careers work*. NICEC Briefing. Cambridge, UK: National Institute for Careers Education and Counselling.

Oravec, J. A. (1996). *Virtual individuals, virtual groups: Human dimensions of groupware and computer networking*. Cambridge, UK: Cambridge University Press.

Reardon, R. C., & Lenz, J. G. (In press). *The Self-Directed Search: A practitioner's guide to the SDS and other Holland-based career materials*. Odessa, FL: Psychological Assessment Resources.

Reardon, R. C., & PAR Staff. (1996). *Self-Directed Search Form R computer version* [Computer program]. Odessa, FL: Psychological Assessment Resources.

Roid, G. H., & Gorsuch, R. L. (1984). Development and clinical use of test-interpretive programs on microcomputers. In M. D. Schwartz (Ed.), *Using computers in clinical practice: Psychotherapy and mental health applications* (pp. 141–149). New York: Haworth Press.

Sampson, J. P., Jr. (1983). Computer-assisted testing and assessment: Current status and implications for the future. *Measurement and Evaluation in Guidance, 15*, 293–299.

Sampson, J. P., Jr. (1990). Computer-assisted testing and the goals of counseling psychology. *The Counseling Psychologist, 18,* 227–239.

Sampson, J. P., Jr. (In press, a). Computer applications. In C. E. Watkins Jr. & V. L. Campbell (Eds.), *Using tests and assessment procedures in counseling* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sampson, J. P., Jr. (In press, b). The Internet as a potential force for social change. In C. C. Lee & G. R. Walz (Eds.). *Social action: A mandate for counselors.* Greensboro: University of North Carolina at Greensboro, ERIC Clearinghouse on Counseling and Student Services.

Sampson, J. P., Jr., Kolodinsky, R. W., & Greeno, B. P. (1997). Counseling on the information highway: Future possibilities and potential problems. *Journal of Counseling and Development, 75,* 203–212.

Acknowledgments

Appreciation is expressed to Thomas Clawson, Janet Lenz, Robert Reardon, and Sandra Sampson for their comments on initial drafts of this paper.

About the Author

Jim Sampson is currently a professor in the Department of Human Services and Studies at Florida State University, where he has taught courses in career development and computer applications in counseling since 1982. Since 1986 he has served as co-director of the Center for the Study of Technology in Counseling and Career Development, a research center established at FSU to improve the design and use of computer applications in counseling and guidance. He writes and speaks on the appropriate use of computer technology in counseling and on the use of cognitive strategies in the improvement of career counseling and guidance services. Prior to joining the faculty at Florida State University, he was a senior counselor and the coordinator of the Career Planning Center at the Student Counseling and Career Planning Center, Georgia Institute of Technology. He is a National Certified Counselor and a National Certified Career Counselor. He received his Ph.D. in counselor education from the University of Florida in 1977.

1. This article was originally the plenary address presented at the Assessment '98: Assessment for Change—Changes in Assessment conference, St. Petersburg, FL, January 16–18, 1998.

Chapter Eighteen

Use of the Kaufman Adolescent and Adult Intelligence Test (KAIT) in the New Millennium

Douglas K. Smith

Abstract

The KAIT is described, with an emphasis on its theoretical base and the distinction between crystallized and fluid intelligence. A synopsis of standardization data as well as reliability and validity data are presented. Most importantly, several uses of the KAIT are described, with two case studies presented to illustrate the usefulness of the test. The KAIT has strong psychometric properties and is well suited for use in both school and clinical settings. The emphasis of the test on the distinction between fluid and crystallized intelligence is a strength, and it contains more subtests measuring fluid intelligence than any other cognitive battery. In addition, all subtests require the use of formal problem-solving skills. The KAIT has a strong theoretical base and offers an additional option for the evaluation of individuals ages 12 to 85 years.

The KAIT, developed by Alan S. and Nadeen L. Kaufman in 1993, is an individually administered test of intelligence for individuals 11 to 85+ years of age. It is a test of general intelligence "composed of separate Crystallized and Fluid Scales. The Crystallized Scale measures acquired concepts and depends on schooling and acculturation for success, while the Fluid Scale measures the ability to solve new problems" (Kaufman & Kaufman, 1993, p. 1). The theoretical base of the test is the result of an integration of Horn and Cattell's theory of fluid and crystallized intelligence; the Luria Golden definition of planning ability; and Piaget's stage of formal operations.

Crystallized and Fluid Intelligence

Crystallized intelligence emphasizes verbal concepts and is heavily influenced by formal school learning. On other intelligence tests, this construct is referred to as verbal intelligence, verbal comprehension, or verbal reasoning. Fluid intelligence emphasizes the ability to solve novel problems. Although this construct is less frequently measured on intelligence tests, examples include Matrix Reasoning on the Wechsler Adult Intelligence Scale third edition (WAIS–III), the Nonverbal Reasoning Cluster on the Differential Ability Scales (DAS), and the Fluid Reasoning Factor on the Woodcock-Johnson–III (WJ–III). In the case of the DAS and WJ–III, there are two subtests composing the cluster and factor, respectively.

Research on factor theories of intelligence, most notably studies by Carroll (1993), Horn (1991, 1994), and McGrew (1997), indicate that in addition to a general factor on intelligence, or *g*, there are additional factors that are key components of *g*. Although there is a lack of agreement on the number of these factors or their names, there is consensus that three factors (crystallized intelligence, fluid intelligence, and visual/spatial intelligence) are the most highly related to *g*. Two of these factors, crystallized intelligence and visual/spatial intelligence, have been the primary emphases of intelligence and cognitive ability tests such as the Wechsler scales and the Stanford-Binet. Recently, however, fluid intelligence has received increased attention with the release of the DAS, the KAIT, the WAIS–III with its new Matrix Reasoning subtest, and the WJ–III. Although all four tests measure fluid intelligence, the KAIT provides the most extensive measure, with four subtests as compared to two subtests in the DAS, one subtest in the WAIS–III, and two subtests in the WJ–III.

Why is fluid intelligence important? The simple answer is that it relates highly to overall intelligence as measured by *g*. Second, it involves a number of important processes related to cognitive skills, including the abilities to reason, solve problems, and form concepts. As Kaufman and Kaufman (1993, p. 11) indicate, "Fluid intelligence (Gf), sometimes called broad reasoning, is the ability to solve new problems, specifically the type that are not made easier by extended education or intensive acculturation." Third, by de-emphasizing acculturation and formal educational experiences, fluid intelligence may be a more appropriate or purer measure of cognitive ability for some individuals.

Structure of the KAIT

The KAIT produces a Composite Intelligence Scale score, a Crystallized IQ score, and a Fluid IQ score, with a mean of 100 and standard deviations of 15. The core battery, administered in one hour, consists of three crystallized subtests (Definitions, Auditory Comprehension, and Double Meanings) and three fluid subtests (Rebus

Learning, Logical Steps, and Mystery Codes). The expanded battery, requiring an additional 30 minutes, consists of the core battery plus two measures of delayed memory (Rebus Delayed Recall and Auditory Delayed Recall) and two alternate subtests, Memory for Block Designs (Fluid) and Famous Faces (Crystallized). The third part of the KAIT is the Mental Status Exam, an optional subtest used when there are concerns as to whether the examinee has the necessary skills to complete the KAIT. The KAIT subtests are described in Table 18.1.

KAIT subtests are organized by the fluid-crystallized distinction and require planning ability and abstract thought. Subtests are primarily either fluid or crystallized, relate to real-life situations, and measure functional skills. Crystallized subtests are presented both verbally and visually with the exception of Auditory Comprehension, which is presented verbally. Fluid subtests are presented visually with verbal directions. The Famous Faces subtest utilizes both a visual and verbal presentation format. The response modality for the KAIT is primarily verbal. The only exceptions are Mystery Codes, in which responses are circled, and Memory for Block Designs, in which the examinee manipulates wooden blocks. Unlike the Wechsler Intelligence Scale for Children (WISC–III) and the WAIS–III, visual-motor coordination and how quickly problems are solved are not emphasized.

Standardization

The KAIT was standardized on 2,000 adolescents and adults from the ages of 11 years to 85+ years, stratified within each age group by gender, geographic region, socioeconomic status (defined by the examinee's or parent's educational level), and race or ethnic group, according to 1988 census data.

Reliability and Validity

Extensive reliability and validity data are presented in the KAIT manual (Kaufman & Kaufman, 1993) and summarized here. Spit-half reliability coefficients for the six core subtests range from .78 to .95, with a mean subtest reliability coefficient of .90 (range of .87 to .93). Average reliabilities for the two alternate subtests are .79 for Memory for Block Designs (range of .76 to .85) and .92 for Famous Faces (range from .83 to .97). Reliability coefficients average .95 for the Crystallized Scale, with a range from .91 to .97; and .95 for the Fluid Scale, with a range from .93 to .96. The average reliability coefficient for the Composite Intelligence score is .97, with a range from .95 to .98. Test-retest reliabilities range from .87 (Fluid) to .94 (Crystallized and Composite).

Table 18.1. Subtests of the KAIT

Subtest	Description
Crystallized Subtests	
Definitions	Examinees figure out a word by studying the word shown with some of its letters missing and hearing or reading a clue about its meaning.
Auditory Comprehension	Examinees listen to a recording of a news story and then answer factual and inferential questions about the story.
Double Meanings	Examinees study two sets of word clues, then think of a word with two meanings that relates closely to both sets of clues.
Famous Faces	Examinees name people of current or historical fame, based on their photographs and a verbal clue about them.
Fluid Subtests	
Rebus Learning	Examinees learn the word or concept associated with a particular rebus (drawing), then "read" phrases and sentences composed of these rebuses.
Logical Steps	Examinees attend to logical premises presented both visually and orally, then respond to a question by making use of the logical premises.
Mystery Codes	Examinees study the identifying codes associated with a set of pictorial stimuli, then figure out the code for a novel pictorial stimulus.
Memory for Block Designs	Examinees study a printed design that is exposed briefly, then copy the design from memory using six yellow and black wooden blocks and a tray.
Delayed Recall Subtests	
Rebus Delayed Recall	Examinees "read" phrases and sentences composed of rebuses they learned about 45 minutes earlier during the Rebus Learning subtest.
Auditory Delayed Recall	Examinees answer literal and inferential questions about new stories they heard approximately 45 minutes earlier during the Auditory Comprehension subtest.

Note: Adapted from Kaufman and Kaufman (1993).

Concurrent validity studies comparing performance on the KAIT with performance on the WISC-R, WAIS-R, and Stanford-Binet fourth edition produced correlations ranging from .57 to .88. Both exploratory and confirmatory factor analyses were performed on KAIT standardization data and are supportive of the factor structure of the test. Since the publication of the KAIT, there have been a number of studies examining its validity (Kaufman & Horn, 1996; Kaufman, Kaufman, & McLean, 1995; Kaufman, McLean, & Kaufman, 1995; Kaufman, McLean, Kaufman, & Kaufman, 1994). These studies have been generally supportive of its validity.

Uses of the KAIT

The purpose of this section is to describe some of the uses of the KAIT within the educational setting. It is especially useful in the assessment of memory problems, fluid reasoning, gifted and talented evaluations, and special education re-evaluations.

Memory Problems

An important feature of the KAIT is the ability to measure both immediate and delayed memory using the Auditory Comprehension and Rebus Learning subtests for immediate memory and Auditory Delayed Recall and Rebus Delayed Recall for delayed memory. The Auditory Comprehension subtests focus on auditory memory, whereas the Rebus Learning subtests focus on visual memory. Norms are provided for determining whether the delayed and immediate versions differ significantly from each other. In interpreting results, I also compare the scores between the two subtests at each level of memory (immediate/delayed) to determine whether there are consistencies or inconsistencies. For some individuals, auditory memory may be significantly better developed for immediate memory (Auditory Comprehension > Rebus Learning) and less well developed for delayed or more long-term memory (Auditory Delayed Recall < Rebus Delayed Recall). It seems reasonable to conclude that most individuals show a basic consistency between the two conditions, and my clinical experience suggests this is the case, although empirical data are lacking.

What is the value of this information? First, it tells us how the individual most effectively takes in information for later retrieval (immediate short-term memory or long-term memory). Is the individual more likely to remember material presented verbally or visually, or does it matter? Secondly, in an academic context it enables us to present information in the most efficient manner for the particular student. The memory data may also be useful in cases where there are changes in the efficiency of either immediate or delayed memory or both over repeated evaluations. Such changes could be the result of aging, neurological difficulties, or accidents to name a few possible causes.

Fluid Reasoning

Of the published tests that measure fluid reasoning or fluid intelligence, the KAIT provides the most extensive measure, with four subtests (see Table 18.2). Therefore, the instrument, especially the fluid subtests, is an important supplement to other cognitive batteries. In fact, it may be the instrument of choice for this reason, with subtests from other batteries being used as supplements. The fluid scale has strong psychometric properties, including an adequate floor and ceiling for ages 11 years, 0 months to 59 years, 11 months. For ages 60 years and older, the floor is less robust with minimum scores based on raw scores of 0 ranging from 53 to 70. See Table 18.3 for the effective range of standard scores across the age range for the Crystallized, Fluid, and Composite scales.

Table 18.2. Fluid Subtests on Various Cognitive Batteries

Cognitive Battery	Fluid Subtests
Differential Ability Scales	Matrices Sequential and Quantitative Reasoning
Kaufman Adolescent and Adult Intelligence Test	Rebus Learning Logical Steps Mystery Codes Memory for Block Designs (optional)
Wechsler Adult Intelligence Scale–III	Matrix Reasoning
Woodcock-Johnson–III	Concept Formation Analysis-Synthesis

Gifted and Talented Evaluations

The KAIT is especially well suited for use with students who may gifted and talented. The ceiling of the test for the three scales (Crystallized, Fluid, Composite) is excellent, as shown in Table 18.3. Second, it is the only cognitive abilities test specifically designed to measure higher-level cognitive processes in the adolescent and adult age ranges. Other instruments are extensions of school-age tests (e.g., Stanford-Binet fourth edition and DAS). Even the WAIS–III is a modification and revision of the WISC–III and the original Wechsler

Table 18.3. Effective Range of Standard Scores

Age	Crystallized IQ	Fluid IQ	Composite Intelligence Scale
11-0 – 11-3	62–160	40–157	48–160
11-4 – 11-7	62–160	40–157	48–160
11-8 – 11-11	59–160	40–157	46–160
12-0 – 12-3	59–160	40–157	46–160
12-4 – 12-7	56–160	40–157	42–160
12-8 – 12-11	53–160	40–157	40–160
13-0 – 13-3	53–160	40–154	40–160
13-4 – 13-7	53–160	40–154	40–160
13-8 – 13-11	53–160	40–151	40–160
14-0 – 14-3	53–160	40–151	40–160
14-4 – 14-7	53–160	40–151	40–160
14-8 – 14-11	53–160	40–151	40–160
15-0 – 15-3	53–160	40–151	40–160
15-4 – 15-7	49–160	40–151	40–160
15-8 – 15-11	49–160	40–151	40–160
16-0 – 16-3	49–160	40–151	40–160
16-4 – 16-7	49–160	40–151	40–160
16-8 – 16-11	45–160	40–151	40–160
17-0 – 17-5	45–160	40–151	40–160
17-6 – 17-11	45–160	40–146	40–160
18-0 – 18-5	41–160	40–146	40–160
18-6 – 18-11	41–160	40–146	40–160
19-0 – 19-11	41–160	40–146	40–160
20-0 – 20-11	41–160	40–146	40–160
21-0 – 22-11	40–160	40–146	40–160
23-0 – 24-11	40–160	40–146	40–160
25-0 – 29-11	40–160	40–146	40–160
30-0 – 34-11	41–160	40–146	40–160
35-0 – 44-11	45–160	40–151	40–160
45-0 – 54-11	53–160	40–151	42–160
55-0 – 59-11	53–160	40–157	44–160
60-0 – 64-11	56–160	53–157	52–160
65-0 – 69-11	62–160	53–157	55–160
70-0 – 74-11	64–160	53–157	57–160
75-0 – 79-11	64–160	60–157	60–160
80-0 – 84-11	69–160	65–157	65–160
85-0+	71–160	70–157	69–160

Note: Based on minimum raw scores (0 on all subtests) and maximum raw scores at each age level.

Bellevue Scale, which originated from David Wechsler's clinical perspectives and experiences at Bellevue Hospital in New York. Third, all of the subtests on the KAIT involve problem-solving skills utilizing what Piaget has described as formal operations. For example, Definitions is a vocabulary test, but the examinee must integrate a visual cue (the configuration of the word) with the word's definition or characteristics in order to produce the correct response. This is a far more complex task than simply defining words that are presented.

The fluid scale may also be useful in identifying gifted and talented individuals who are missed by programs that emphasize crystallized intelligence and academic achievement. Traditional measures of cognitive ability have emphasized verbal intelligence rather than fluid intelligence. Thus, individuals skilled in nonverbal problem solving or solving novel problems are often not identified as gifted and talented, even though their skills in this aspect of cognitive ability may be quite well developed.

Special Education Re-evaluations

The IDEA '97 amendments provide increased flexibility in special education re-evaluations. For example, the disability does not need to be rediagnosed. Emphasis is placed on obtaining information that will be useful in educational programming and transition planning. Thus, examiners are relieved of the task of readministering that same cognitive ability evaluation year after year.

By far the most frequently administered test of cognitive abilities in both school and clinical settings is the WISC–III/WAIS–III (Oakland & Hu, 1992; Stinnett, Havey, & Oehler-Stinnett, 1994; Watkins, Campbell, Nieberding, & Hallmark, 1996). All too often re-evaluations have simply consisted of readministering the same test without adding any new information. In my experience as a school psychologist and a trainer of school psychologists, it is typical to see the WISC–III, and at older ages the WAIS–III, administered to special education students three, four, or even five times. Each time the scores and profiles are similar and little new information is added.

With the new IDEA '97 amendments, the examiner can supplement the evaluation with additional information. The KAIT provides the opportunity to provide information on fluid reasoning and both immediate and delayed memory, for example. Although empirical data are currently lacking, clinical experiences indicate that in some instances students diagnosed as having cognitive or learning disabilities have shown relative strengths in fluid reasoning upon re-evaluation with the KAIT. For example, Donald, age 16 years, 3 months, was originally diagnosed as having a cognitive disability in the second grade. At age 15 he was re-evaluated with the WISC–III and the KAIT.

His WISC–III scores were consistent with previous scores and included a Verbal IQ of 69, a Performance IQ of 59, and a Full Scale IQ of 61. These scores indicated that Donald was functioning at a level well below his peers. On the KAIT, however, he obtained the following scores: Crystallized IQ of 83, Fluid IQ of 77, and Composite IQ of 79. These scores are still below average, but they do suggest a somewhat higher potential than previously indicated. In contrast to his performance on the WISC–III, in which only Digit Span was in the average range, Donald had scores in the average range on three KAIT subtests (Double Meanings, Famous Faces, and Logical Steps). These scores suggest that his problem-solving skills, although below average, show more potential than previously indicated. In addition, the KAIT results suggest that transition activities for Donald should focus on developing vocational skills that would allow him to secure employment and become self-sufficient.

Another case is Jeremy. He is also 16 years of age (16 years, 11 months) and has been receiving services for a learning disability for the past nine years. At his latest re-evaluation he was administered the WISC–III and the KAIT. On the WISC–III he received a Verbal IQ of 89, a Performance IQ of 97, and a Full Scale IQ of 97, whereas on the KAIT his scores were 94 for Crystallized IQ, 122 for Fluid IQ, and 108 for Composite IQ. Once again the measures of verbal ability (crystallized ability) are similar, but a strength both relative and in relation to peers emerges in fluid reasoning. More important, Jeremy displayed large discrepancies between his scores in the measures of auditory memory and visual memory. His mean auditory memory score was 10.5 (Auditory Comprehension = 6, Auditory Comprehension Delayed = 5) and the mean visual memory score was 14.5 (Rebus Learning = 14, Rebus Learning Delayed = 15). These results suggest that he excels in solving novel nonverbal problems and that he is most likely to remember and retrieve information that presented visually rather than orally.

Although these two cases do not constitute empirical data, they do show the utility of the KAIT on a case-by-case basis. With now flexibility in re-evaluation procedures, it is feasible to collect additional information that may prove useful in programming and transition planning.

References

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock, *WJ-R technical manual*. Chicago: Riverside Publishing.

Horn, J. L. (1994). Theory of fluid and crystallized intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 443–451). New York: Macmillan.

Kaufman, A. S., & Horn, J. L. (1996). Age changes on tests of fluid and crystallized ability for females and males on the Kaufman Adolescent and Adult Intelligence Test at ages 17 to 94. *Archives of Clinical Neuropsychology*, *11*, 97–121.

Kaufman, A. S., Kaufman, J. C., & McLean, J. E. (1995). Factor structure of the Kaufman Adolescent and Adult Intelligence Test (KAIT) for Whites, African-Americans, and Hispanics. *Educational and Psychological Measurement*, *55*, 365–376.

Kaufman, A. S., & Kaufman, N. L. (1993). *Kaufman Adolescent and Adult Intelligence Test*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., McLean, J. E., & Kaufman, J. C. (1995). The fluid and crystallized abilities of white, black and Hispanic adolescents and adults, both with and without an education covariate. *Journal of Clinical Psychology*, *51*, 637–647.

Kaufman, J. C., McLean, J. E., Kaufman, A. S., & Kaufman, N. L. (1994). White-Black and White-Hispanic differences on fluid and crystallized abilities by age across the 11- to 94-year range. *Psychological Reports*, *75*, 1279–1288.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In Flanagan, D. P., Genshaft, J. L., & Harrison, P. L. (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151–179). New York: Guilford.

Oakland, T., & Hu, S. (1992). The top 10 tests used with children and youth worldwide. *Bulletin of the International Test Commission*, *19*, 99–120.

Stinnett, T. A., Havey, J. M., & Oehler-Stinnett, J. (1994). Current test usage by practicing school psychologists: A national survey. *Journal of Psychoeducational Assessment*, *12*, 331–350.

Watkins, C. E., Campbell, V. I., Nieberding, R., & Hallmark, R. (1996). On Hunsley, harangue, and hoopla: Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice*, *27*, 316–318.

About the Author

Douglas K. Smith is currently director of programs in school psychology at the University at Albany–State University of New York. He obtained his Ph.D., Ed.S., and M.Ed. degrees in school psychology from Georgia State University. Current research interests include psychoeducational assessment issues in general and developing individual testing accommodations for students with disabilities in particular. Dr. Smith is author of *Essentials of Individual Achievement Assessment* (2001) and co-editor of the forthcoming *Assessing People With Disabilities in Educational, Employment, and Counseling Settings*, as well as numerous journal and chapter articles. Dr. Smith was named the Outstanding Faculty Member of 1987 in the College of Education at the University of Wisconsin–River Falls.

Chapter Nineteen

Writing Multiple-Choice Test Items

Nicholas A. Vacc, Larry C. Loesch, & Ruth E. Lubik

Abstract

Multiple-choice tests are widely viewed as the most efficient and objective means of assessment. Item development is the most critical component of creating an effective test, but unfortunately, most test developers have no background in item development. The three cognitive levels of test items (recall, application, and analysis) are described, along with the three main item types (single best response, situational set, and complex). Finally, guidelines for writing appropriate and effective item stems, keyed responses, and distracters are provided.

Most adults have taken a multiple-choice test at some time in their lives. Such tests frequently are used in educational systems to assess academic aptitude or achievement, and they frequently are used in job application processes to determine an applicant's potential or skills. They also often are used in professions as part of a licensure or certification application process (Karras, 1991; Vacc, 1991). Clearly, tests are viewed by many as the best and most efficient way to gather and evaluate data and information.

Because multiple-choice tests are used widely and because they have significant impact on the lives of those taking them, using procedures that are proven effective for their development is important. Cohen and Swerdlik (1999, p. 215) indicated, "The creation of a good test is not a matter of chance—it is the product of the thoughtful and sound application of established principles of test construction." Such principles are found in resources such as the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1985), *Responsibilities of Users of Standardized Tests* (AACD & AMECD, 1989), and *Code of Fair Testing Practices* (Joint Committee on Testing Practices, 1988). Each set of principles has as its goal the development of an instrument that has a high level of objectivity and validity, because well-produced tests increase the likelihood that test scores can be of assistance (Vacc, 1991).

Haladyna and Downing (1989) noted that one of the most important steps in test development is item writing. They concluded that test quality therefore is contingent upon the quality of test items. Unfortunately, McDougall (1997) and Osterlind (1989) stated that most test developers construct tests based on "folk wisdom" rather than a systematic application of principles of effective item development. Most likely, the lack of a systematic procedure occurs because few professionals are trained adequately in test construction; therefore, they focus on test information interesting to themselves rather than on essential material. The unfortunate result often is item-writer bias (Haladyna, 1992; McDougall, 1997). Even highly educated college faculty typically lack effective test-development training and thus make similar errors (McDougall, 1997).

Despite common and widespread problems in test construction, multiple-choice tests remain popular and appear to be dominant among objective tests (Haladyna, 1992; Haladyna & Downing, 1989; McDougall, 1997; Pomplun & Omar, 1997). Multiple-choice tests afford fast, relatively accurate, economical, and objective ways to obtain data, and they have the advantage of being applicable to a wide range of topics (Cohen & Swerdlik, 1999). Multiple-choice tests also are generally thought to be reliable, versatile, and easily used (Haladyna & Downing, 1989; Karras, 1991; McDougall, 1997).

Haladyna (1992) suggested that better measurement of both achievement and abilities could be achieved most easily through improvements in item writing. Haladyna and Downing (1989, p. 47) compiled 43 item-writing guidelines, rules, and suggestions from various textbooks, and concluded that applying these guidelines would result in tests that are uniform in appearance and free of nettlesome item-writing faults and other problems that distract examinees from giving their best responses.

Most multiple-choice items can be classified into one of three cognition levels: recall, application, and analysis. Each level utilizes a different cognitive function:

Recall-level items: Recall-level items primarily test the recognition or recall of relatively isolated facts, concepts, principles, processes, procedures, or theories. Responding correctly to items at this level is primarily a function of an individual's memory. Incorrect responses result when the individual is unable to remember or recall the answer.

Application-level items: Application-level items primarily test relatively simple interpretations or limited applications of data or information. Items at this level require more than application of memory; responding correctly requires relatively minor or low-level problem-solving skills.

Analysis-level items: The third commonly used level of multiple-choice items is the analysis level. Items at this level primarily test skills involving evaluation of data, problem solving, or the fitting together of elements into a meaningful whole. Responding correctly to these items involves application of both good judgment and problem-solving skills. This level thus involves higher cognitive processes than the other levels.

Item Types

Multiple-choice items also can be classified by type, with each type having unique characteristics and challenging a respondent's thinking in different ways. Three commonly used types of multiple-choice items are single best response, situational set, and complex.

Single best response items: The most commonly used type is the single best response item. With this type of item, there purportedly is one correct answer among the various response choices (sometimes called the distracters or foils) for the item. Single best response items may be developed in several forms. One form is the direct question in the item stem to which the respondent is required to provide the answer from the response choices. Another form is an incomplete statement in the item stem for which the respondent is asked to select the word or phrase from among the choices that best completes it. The third form is the calculation item for which the respondent is required to perform some calculation, usually mathematical, in order to determine the correct response from among the choices.

Situational set items: The situational set item presents a scenario containing a collection of facts or data, followed by the item stem. Typically, there are three to five multiple choices associated with each situational set, usually of the single best response form. However, each choice is expected to stand alone and is not contingent upon any other for correct responding.

Complex items: The complex item requires simultaneous consideration of several facts or bits of information. A complex item consists of a stem followed by three to five statements, phrases, or sometimes graphic depictions known as the elements. The distracters in the item include combinations of the elements. Respondents to these types of items face an all-or-none dilemma; knowing only one of the elements will not allow determination of the correct response.

Writing Multiple-Choice Items

Theoretically, the correct way to respond to a multiple-choice test question is not by eliminating the incorrect responses and then choosing from the remaining responses, but rather by reading the item stem carefully, formulating the correct response based on the information in the item stem, and then finding the correct response from among the distracters. The approach to responding has significant implications for writing effective multiple-choice items. For example, the item stem must be written so that respondents can formulate the correct response mentally before considering the distracters. In addition, effective distracters are created through consideration of how respondents might think incorrectly or illogically in responding to the item stem.

Writing Item Stems

There are several guidelines to follow in constructing item stems effectively and efficiently. One is to use clear and simple language. The use of jargon and highly technical vocabulary should be avoided unless they are appropriate for the purpose of the item. An item developer also should use simple sentences and grammatical constructions that promote ease of reading and understanding for the respondent.

A second guideline in stem construction is to present only a single, clearly formulated idea or problem. Item developers should avoid including multiple ideas or vague or ambiguous concepts in the item stem. In addition, test items should focus on general knowledge and principles and be devoid of unnecessary specificity; excessive "window dressing" or irrelevant information defeats the goal of effective assessment.

The last major item-stem development guideline is to put as much of the wording as possible in the stem rather than writing a short item stem with numerous distracters. In fact, all the information or qualifications necessary to determine the correct answer should be in the item stem. At the same time, however, item developers should avoid using a literal definition as the item stem. Rather, the stem should provide the information in clear, easily understood language. Finally, the use of negative wording (e.g., "which of the following is not") should be avoided as much as possible.

Writing Distracters

Formulating distracters with care is important so that irrelevant characteristics do not trigger responding behaviors. Foremost, an item developer must insure that the keyed response (i.e., the one to be scored as correct) is both correct and clearly the best response. The distracters in a multiple-choice item should be independent of one another, arranged in logical order, and grammatically consistent with the stem. They also should not cue responding to answers or distracters in other items. In general, item developers should avoid using phrases such as "all of the above" or "none of the above" as distracters.

Multiple-choice item distracters should be designed to be attractive to respondents who do not have a good understanding of the content of the item stem. One reasonably effective method of constructing such distracters is to use common misconceptions about the content in the item stem. Using "good-sounding" words in the distracters, such as *accurate, important,* or *significant* often is effective. Also, good distracters should be similar to the keyed response in length, complexity, and grammatical structure. Presenting distracters in language familiar to respondents and avoiding distracters that contradict each other are other effective strategies.

General Guidelines for Test Items

A test developer must decide upon the most effective and efficient format possible for testing the desired material. Irrelevant sources of difficulty should be avoided, as should items that cue responses for other items. Normal and correct rules of grammar and spelling should be used and the use of gender-specific pronouns should be avoided.

If the stem is a question, each distracter should begin with a capital letter and end with a period because the distracters are not continuations of the item stem. When the item stem is an incomplete sentence, each distracter should begin with a lower-case letter. Periods should be omitted following numeric distracters to avoid confusion with decimal points.

Irrelevant clues to the keyed response should be avoided by having essentially similar language in the stem and the keyed response and by avoiding buzzwords that give away the keyed response. Additionally, vague modifiers, such as *sometimes, usually,* or *may,* should be avoided, as should absolute terms such as *always, never, none,* or *only.* Essentially equivalent distracters should also be avoided.

Other important concerns in effective item development are to keep the reading level of the item stem and distracters as low as possible, and to avoid the repetitive use of favorite phrases, terms, or grammatical constructions. Items or questions for which the correct response is merely an opinion also should be avoided unless the source of the opinion is identified clearly. Item content tied to a specific reference, such as a textbook or journal article, should be avoided unless a particular perspective is being espoused, in which case the source must be identified clearly.

It is good psychometric practice to have items reviewed for clarity and cogency before their initial administration, preferably by persons similar to the intended respondents. Item performance characteristics also need to be examined after each administration, particularly those relative to item difficulty, discrimination, reliability, and validity. In effect, each item is field tested in each administration by reviewing the results and item data, and revising as appropriate.

Conclusion

Knowing how to construct good multiple-choice items has important implications for counselors. Indeed, the codes of ethics of the American Counseling Association and the National Board for Certified Counselors call for professional counselors to be knowledgeable of testing and test construction. These admonitions are made because counselors frequently are involved in test use and evaluation, either as test users or test developers, and they frequently help develop tests that are used to evaluate other individuals. In addition, important and significant judgments about individuals and programs are made based on test scores. Thus, if counselors are to fulfill their professional functions and obligations effectively and fully, they must be knowledgeable in effective test- and item-development practices.

References

American Association for Counseling and Development (AACD) & Association for Measurement and Evaluation in Counseling and Development (AMECD). (1989). *Standards for users of standardized tests.* Alexandria, VA: AACD.

American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). (1985). *Standards for educational and psychological testing.* Washington, DC: APA.

Cohen, R. J., & Swerdlik, M. E. (1999). *Psychological testing and assessment: An introduction to tests and measurement* (4th ed.). Mountain View, CA: Mayfield.

Haladyna, T. M. (1992). The effectiveness of several multiple-choice formats. *Applied Measurement in Education, 5*, 73–88.

Haladyna, T. M., & Downing, S. M. (1989). A taxonomy of multiple-choice item writing rules. *Applied Measurement in Education, 2*, 37–50.

Joint Committee on Testing Practices. (1988). *Code of fair testing practices*. Washington, DC: APA.

Karras, R. W. (1991). Let's improve multiple-choice tests. *OAH Magazine of History, 6*, 8–9.

McDougall, D. (1997). College faculty's use of objective tests: State-of-the practice versus state-of-the-art. *Journal of Research and Development in Education, 30*, 183–193.

Osterlind, S. J. (1989). *Constructing test items*. Boston, MA: Kluwer Academic.

Pomplun, M., & Omar, M. H. (1997). Multiple-mark items: An alternative objective item format. *Educational and Psychological Measurement, 57*, 949–962.

Vacc, N. A. (1991). Changing times: Changing views about testing. *Journal of Humanistic Education and Development, 30*, 88–90

About the Authors

Nicholas A. Vacc is the Joe Rosenthal Excellence Professor and past chairperson of the Department of Counseling and Educational Development at the University of North Carolina at Greensboro. In 1991 and 1994, during his tenure as chair, the department was selected as the Outstanding Counselor Education Program by the Association for Counselor Education and Supervision. He has served as president of the Association for Assessment in Counseling, editor of *Measurement and Evaluation in Counseling and Development*, and president of Chi Sigma Iota International. He also is a member of numerous technical and policy advisory boards and working groups concerned with assessment in counseling and mental health, including the Joint Committee on Testing Practices.

Larry C. Loesch is currently a professor in the Department of Counselor Education at the University of Florida, where he has taught since he received his Ph.D. from Kent State University in 1973. A national certified counselor, his teaching interests include research, measurement and evaluation, and supervision of students' research and counseling. Loesch served as president of both the Florida and national Associations for Measurement and Evaluation in Guidance, the Florida Counseling Association, SACES, and Chi Sigma Iota. He was a corecipient of the American Counseling Association's 1983 Research Award and of its 1992 Arthur A. Hitchcock Distinguished Professional Service Award. Loesch was a Fulbright Scholar in Slovakia in 2001.

Ruth E. Lubik is a substance abuse counselor in the Randolph County, North Carolina Mental Health Department. She is a graduate of the University of North Carolina at Greensboro with a M.S. and an Ed.S. from the Counseling and Educational Development Department.

Assessment in Counseling and Therapy

An ERIC/CASS Special Digest Collection

William D. Schafer, Editor

Table of Contents

ASSESSMENT IN COUNSELING & THERAPY SPECIAL DIGEST COLLECTION

Overview of Assessment in Counseling and Therapy *229*
William D. Schafer

Assessment Skills of Counselors, Principals, and Teachers *235*
James C. Impara

Assessment Skills for School Counselors *241*
William D. Schafer

Mental Health Counseling Assessment: Broadening One's Understanding of the Client and the Client's Presenting Concerns *247*
Gerald A. Juhnke

CACREP Accreditation: Assessment and Evaluation in the Standards and Process *253*
Carol L. Bobby & Joseph R. Kandor

The Role of Assessment in Counselor Certification *259*
Thomas Clawson

Assessment of Counselor Performance *265*
Larry C. Loesch

Evaluating School Guidance Programs *271*
Norman C. Gysbers

New Assessment Methods for School Counselors *277*
W. James Popham

Developing and Using Sound Performance Assessments in the Guidance Context *281*
Richard J. Stiggins

Portfolios for Assessment and Instruction *287*
Judith A. Arter, Vicki Spandel & Ruth Culham

Emerging Student Assessment Systems for School Reform *293*
Edward Roeber

Assessment of Abilities *299*
Thomas F. Harrington

Interest Assessment *305*
Jo-Ida C. Hansen

Assessment of Self-Concept *309*
William Strein

Assessment of Temperament *313*
Hedwig Teglasi

Assessment of Preschool Children *319*
Nicholas A. Vacc & Sandra H. Ritter

Screening for Special Diagnoses *325*
Susan De La Paz & Steve Graham

Assessment in Career Counseling *329*
Dale J. Prediger

Assessing Career Certainty and Choice Status *335*
Paul J. Hartung

Gender Differences in Adolescent Career Exploration *339*
Helen S. Farmer

Assessing Employability Skills *345*
Thomas H. Saterfiel & Joyce R. McLarty

Assessing Career Development With Portfolios *349*
Juliette N. Lester & Nancy S. Perry

Ethics in Assessment *355*
Cynthia B. Schmeiser

Multicultural Assessment *361*
William E. Sedlacek & Sue H. Kim

Fairness in Performance Assessment *365*
Tony C. M. Lam

Computer-Assisted Testing in Counseling and Therapy . *371*
James P. Sampson, Jr.

Testing Students With Disabilities *375*
Kurt R Geisinger & Janet F. Carlson

The School Psychologist's Role in School Assessment *381*
Sylvia Rosenfield & Deborah Nelson

Cooperation Between School Psychologists and Counselors in Assessment *387*
Douglas K. Smith

Using Buros Institute of Mental Measurements Materials in Counseling and Therapy *393*
Barbara S. Plake & Jane Close Conoley

Internet Resources for Guidance Personnel *399*
Liselle Drake & Lawrence M. Rudner

Locating and Evaluating Career Assessment Instruments *405*
Jerome T. Kapes

Inappropriate Statistical Practices in Counseling Research: Three Pointers for Readers of Research Literature *411*
Bruce Thompson

Overview of Assessment in Counseling and Therapy

William D. Schafer

The wide range of individual differences among persons with whom counselors deal on a day-to-day basis places extraordinary demands on counselors to understand their clients and their clients' concerns throughout all phases of their work. Assessments are fundamental to those understandings. That there exists a wealth of literature on a wide scope of topics about assessment in counseling and therapy should not be surprising. The purpose of this digest series is to summarize major portions of that literature. Recognized professionals have written about topics that are grouped here into nine broad areas. This foreword gives an overview of the authors' explorations of those topics.

(1) Assessment in Counselor Education and Evaluation

Counselors' understandings about assessment are fundamental to effective use of those techniques. Focusing on school settings, Impara compares assessment knowledge of school counselors, principals, and teachers with each other and with existing standards, concluding that all three groups show uneven skills across important assessment topics. The historical debate over school counselors' uses of assessments is reviewed by Schafer, who also describes needed skills based on a review of job analyses of school counselors. These skills are associated with three roles: pupil assessment, program evaluation, and using basic research. Juhnke considers mental health counselors' uses of assessment. He describes some assessment techniques that can be used along with testing, including qualitative assessment approaches, behavioral assessments, and use of past records. He concludes that combinations of assessment methods, used continuously, will best promote effective treatment strategies.

There are two levels of certification for counselors. One is certification of counselor education programs. Bobby and Kandor describe the process used by the Council for Accreditation of Counseling and Related Educational Programs (CACREP) to compare programs in higher education with existing standards. The other is individual certification of counselors. The strengths and weaknesses of various types of information used in voluntary certification processes are summarized by Clawson. He suggests ways in which the National Board for Certified Counselors (NBCC) may modify its methods to incorporate new data sources in its national certification program.

Delivery of service should be evaluated for both formative and summative purposes. Again, there are two levels of evaluation. Assessment of the performance of individual counselors is discussed by Loesch, who considers multiple data sources and problems in the areas of validity and generalizability associated with them. Gysbers describes ways in which school guidance programs are being evaluated using information from three areas: program standards, job descriptions, and results.

(2) New Forms of Assessment

At both individual and institutional levels, assessment techniques are being used in new ways with greater and greater frequency and are being evaluated against new criteria, such as the consequences of their use for educational and other programs. Popham describes some of these new forms of assessment and considers roles counselors can play in order to use and help others use them more effectively.

There are two basic types of assessment that are receiving more emphasis. One of these is performance assessment. Stiggins summarizes how to develop and evaluate performance assessments and relates this form of assessment to the roles of counselors. The other is portfolio assessment. Arter, Spandel, and Culham describe the many uses of portfolios, how those uses can be made most effective, and some issues that need to be resolved about their use, particularly in high-stakes applications. Roeber relates new systems of assessment to the school reform movement at the national, state, and local levels. He challenges those involved in assessment policy to coordinate their efforts both to assist schools and to document their effectiveness.

(3) Assessment of Traits

Individuals differ from one another in more ways than we will ever be able to assess. The focus here is on abilities, interests, self-concept and temperament. Harrington describes fifteen abilities and considers the use of self-estimates as a promising means to enhance self-awareness. Hansen reviews the historical development of assessments of interests, describes the approaches used in major existing inventories, and explores uses of computers in interest assessment.

While abilities and interests are assessed routinely in most educational settings, self-concept, and temperament are evaluated more often for research purposes or when in-depth understanding of an individual is needed. The nature of self-concept as a trait is considered by Strein, who also describes some commonly used self-concept measures. He offers cautions for counselors to consider in using existing assessments. Teglasi reviews ways in which temperament has been conceptualized and how those conceptualizations have been expressed in existing measures. She also relates temperament to personality and points out the need for refinement of both the construct of temperament and its assessment in educational and mental health contexts.

(4) Assessment for Diagnosis

Assessment plays a crucial role in delivery of special services to individuals. Vacc and Ritter note that all children are entitled by law to appropriate and free educational and other services regardless of their disabilities. They describe trends in assessment for diagnosis at the preschool level and the roles various professionals, including mental health practitioners, play in the process. Once children enter school, screening generally focuses on diagnoses different from those at the preschool level. De La Paz and Graham discuss identification of disabilities in elementary grades through high school. They highlight increased use of better screening procedures and pre-referral interventions as promising ways to address problems of misclassification.

(5) Assessment in Career Development

In career development, both individual and environmental explorations are needed to help individuals lead full, productive lives. Using the premise that a career is a way to implement a self-concept, Prediger attempts to formulate a coherent understanding about the design of assessment components in programs for career development. Hartung focuses on modern approaches to assess career indecision and choice status. Problems of stereotyping by gender are discussed by Farmer, who concludes that counselors bear responsibility for gender-fair career development in order to implement the exploration validity of career interest measurements.

Employability assessment is an area that has received markedly increasing attention in recent years. Saterfiel and McLarty describe assessment of employability skills and discuss several examples that expand on attempts to identify what those skills are. Uses of portfolios to capitalize on the development of understandings about employability in designing and evaluating career development programs are considered by Lester and Perry.

(6) Social Context of Assessment

Throughout all assessment applications, it is important to engage in fair and ethical practices with respect to persons, both individually and in groups. Schmeiser reviews several ethical statements that pertain to assessment, raises issues about their enforcement, and offers some suggestions about including ethics in the education of professionals. Sedlacek and Kim describe ways in which assessments are commonly misused in multicultural settings and how professionals can guard against these misuses, as well as areas of needed research. In the context of performance assessments, Lam differentiates two orientations to fairness: equality and equity. He concludes that each view has both positive and negative ramifications.

(7) Modifications for Special Assessment Circumstances

Special circumstances is used here to describe both measurement conditions and individual needs. Sampson identifies five ways in which computers may be incorporated into assessment and discusses their benefits and limitations. He concludes that counselors should play a major role in shaping applications of computer technology in their fields of practice.

Implications of the Americans with Disabilities Act for assessment are described by Geisinger and Carlson. Most of these are in the areas of test selection, administration, and interpretation. Some practical suggestions for counselors are discussed.

(8) School Psychologists' Roles in Assessment

Rosenfield and Nelson review historical roles of school psychologists in assessment. They differentiate current practices into three areas: making entitlement/classification decisions, planning interventions, and evaluating outcomes. They argue for greater emphasis on collaboration. Echoing that emphasis, Smith describes advantages of collaboration between school psychologists and counselors; broader, multidisciplinary teams are also considered.

(9) Assessment Professionalism

Plake and Conoley describe materials produced by the Buros Institute of Mental Measurements, including publications, symposiums, library collections, CD-ROMs, and a desk reference series for the individual practitioner. Drake and Rudner offer an assessment-oriented tour of the information superhighway. They describe opportunities that are available through selected listservs, gopher sites, and an ERIC e-mail resource and, in a user-friendly way, take us through how to do it.

Kapes describes how to locate and evaluate career assessment instruments. He reminds us that the user is responsible for the final judgment about whether a particular instrument is appropriate.

As we read research and test reviews, we should be aware of the need to be critical consumers of the information. Thompson describes three prevalent inappropriate practices that we should be alert for: ascribing reliability to tests, confusing statistical significance with practical importance, and using stepwise selection of variables in multiple regression (and other) contexts.

Final Note of Appreciation

Many thanks are due the authors of these digests. Active professionals are used to writing tersely, but hardly ever under such a stringent length limitation for such a broad topic as each one of these digests represents. Several agonized phone calls and e-mail messages over the last couple of months attest to both frustration and perseverance on the part of the authors. Meeting stringent deadlines was also necessary. That so many busy professionals accepted and met this challenge is testimony to the spirit of research dissemination that characterizes our profession and is fostered by the ERIC clearinghouses.

It has been a pleasure to serve as a guest editor in the ERIC/CASS digest series program. I hope you, the consumer, feel your time reading these digests is time well spent.

William D. Schafer is Associate Professor of Measurement, Statistics, and Evaluation, University of Maryland, College Park.

Assessment Skills of Counselors, Principals, and Teachers

James C. Impara

There are several methods one might use to determine the level of skills and knowledge of educational practioners in the area of student assessment. One method is to survey various groups of education professionals and ask them to self-report on the extent of their knowledge (or their confidence) in skills associated with student assessment. This is the approach typically taken by researchers who have investigated the topic among counselors (Elmore, Ekstrom, & Diamond, 1993), principals, and teachers (Fennessey, 1982; Infantino, 1976). A second way to undertake research in this area is to develop a test of assessment skills and knowledge and administer it to groups of counselors, principals, and teachers. This approach was used by Impara, Divine, Bruce, Liverman & Gay (1991) and by Impara and Plake (in press). A third method, particularly suitable for teachers, is to examine the tests they develop and infer their knowledge of principles of test construction (Gullickson & Ellwein, 1985); this method provides only limited information about their knowledge of assessment skills.

A precursor to measuring the assessment skills of educational professionals is identifying the skills to be measured. This might be done by undertaking a job analysis, e.g., asking counselors, principals, and teachers what assessment skills and knowledge they need to perform their job. Another way is to seek appropriate professional standards that might define the scope and level of assessment skills and knowledge needed.

Standards for Assessment

The major, and most general, standards are the *Standards for Educational and Psychological Testing*, ([AERA] American Educational Research Association), (]APA] American Psychological Association), & ([NCME]National Council on Measurement in Education, 1985). More directly relevant to assessment skills are the standards that have been (or are being) developed by professional organizations responsible for certifying or otherwise imposing some degree of control or direction over the profession. Among the standards developed for counselors that are relevant to assessment are: Responsibilities of Users of Standardized Tests (American Association for Counseling and Development [AACD]/Association for Measurement and Evaluation in Counseling and Development [AMECD], 1989); *Ethical Standards* [AACD], 1989 (currently under revision); and the CACREP Accreditation Standards (Council for Accreditation of Counseling and Related Educational Programs, 1994).

In a joint endeavor the American Federation of Teachers (AFT), NCME, and the National Education Association (NEA) produced the *Standards for Teacher Competence in Educational Assessment of Students* (1990). In a follow up to that effort the American Association of School Administrators (AASA), National Association of Elementary School Principals (NAESP), National Association of Secondary School Principals NASSP), & NCME have drafted the *Competency Standards in Student Assessment for Educational Administrators*. (these standards should be available from the participating organizations by mid 1995)

The Research Findings on Skills and Knowledge of Educational Professionals

Elmore et al. (1993) surveyed counselors, in part to collect information related to the measurement dimensions of the *Ethical Standards* (AACD, 1988). The questionnaire asked counselors about their level of confidence associated with undertaking various assessment activities. The results indicated that many counselors feel highly confident about using test results (69%), selecting tests (67%), administering tests (90%), and interpreting test scores (72%). Counselors also reported high levels of confidence in using test norms (72%); using statistics like the mean, standard deviation, and correlation (67%); using test reliability and validity information (59%); and using the standard error of measurement (58%) (Elmore et al., 1993, p. 118).

Impara et al. (1991) investigated the extent that elementary and secondary teachers' interpretation of a standardized test score report from a state testing program was aided by the interpretative information provided by the scoring service. They found that teachers who had the interpretive information made fewer errors responding to test questions based on the score report than did teachers who did not have the benefit of interpretive information. (14 of 17 correct vs. 12 of 17 correct) The most difficult items for all the teachers related to interpreting percentile bands. Some teachers, especially those at the secondary level, commented that they did not have to know how to interpret test scores because they could rely on the school counselors to interpret and explain test scores to students.

In a later study, Impara and Plake (in press) obtained responses from over 900 Virginia educators (balanced about equally among counselors, principals, and teachers at both elementary and secondary levels) on a test developed using as test specifications the *Standards for Teacher Competence in Educational Assessment of Students* (AFT, NCME, & NEA, 1990). Counselors' strengths were associated with items relating to test selection, validity, communication of assessment results, and ethical practices. Unlike both principals and teachers, counselors showed particular strength in their basic understanding of the concept of reliability and measurement error, and their ability to interpret scores from standardized tests. In contrast to counselors, both principals and teachers more often confused reliability and validity.

Principals showed strength in understanding the bases for selecting an assessment strategy and the methods for determining validity. Most principals also answered correctly items addressing communication of test results, but (like teachers and counselors) were less proficient in the interpretation of standardized test results. Finally, principals' scores were very high on the items measuring the recognition of ethical practices.

Although teachers' strengths were similar to those identified for principals and counselors, many teachers (about 37%) did not understand the correct interpretation of grade equivalent scores. All respondents had problems understanding how to combine scores from individual assessments, e.g., several tests, into a single summary grade. As in Impara et al. (1991), many teachers, especially those in secondary schools, indicated they rely on counselors to provide interpretations of standardized tests.

In terms of the overall performance of the different levels of professionals in this study, the counselors at both elementary and secondary levels and the elementary principals received higher scores than did either the teachers or secondary principals. It is clear that teachers rely on counselors and that this group of professionals is expected to serve in a consulting role to other professionals within the school in many matters of testing and assessment, especially when dealing with formal testing programs. In elementary schools where counselors are least likely to be available, principals may need to serve in the same consultative capacity as counselors do in high schools, so they, too, must be adequately prepared to assist teachers in matters related to formal testing programs. As a group, however, none of the professionals surveyed is well prepared in the development and use of assessments at the classroom level.

Summary and Conclusions

The findings from Elmore et al. (1993), Impara et al., (1991) and Impara & Plake (in press) parallel each other and those from the self-report studies reported by other researchers in that many educational professionals have some knowledge of assessment practices, ranging from principles of test development and use to the practices associated with the use and interpretation of standardized and teacher-made tests. The skill levels associated with many important student assessment principles is, however, not consistent with the standards adopted by professional organizations.

The various standards that have been developed and endorsed by the professional associations in education are important documents, and they provide excellent guides for the professional development of educators who work with assessment information on a regular basis. Clearly the assessment skills and knowledge of counselors, principals, and teachers are lacking in some important areas while in other important areas these educational professionals are highly skilled and knowledgeable.

References

American Association for Counseling and Development. (1988). *Ethical Standards.* Alexandria, VA: Author.

American Association for Counseling and Development/Association for Measurement and Evaluation in Counseling and Development. (1989, May). Responsibilities of users of standardized tests. *Guidepost*, 12, 16, 18, 27, 28.

American Federation of Teachers, National Council on Measurement in Education and National Education Association. (1990). *Standards for Teacher Competence in Educational Assessment of Students.* Washington, DC: American Federation of Teachers.

American Psychological Association, American Educational Research Association,and National Council on Measurement in Education. (1985). *Standards for Educational and Psychological Testing.* Washington, DC: American Psychological Association.

Council for Accreditation of Counseling and Related Educational Programs. (1994, January). *The CACREP Accreditation Standards and Procedures Manual.* Alexandria, VA. Author.

Elmore, P. B., Ekstrom, R., & Diamond, E. E. (1993). Counselors' test use practices: Indicators of the adequacy of measurement training. *Measurement and Evaluation in Counseling and Development, 26*(2),116-124.

Fennessey, D. (1982, July). *Primary teachers' assessment practices: Some implications for teacher training.* Paper presented at the annual meeting of the South Pacific Assn. for Teacher Education, Frankston, Victoria, Australia. (ERIC Document Reproduction Service No. ED 229346).

Gullickson, A. R., & Ellwein, M. C. (1985). Post hoc analysis of teacher-made tests: The goodness-of-fit between prescription and practice. *Educational Measurement: Issues and Practice, 4* (1), 15-18.

Impara, J. C., Divine, K. P., Bruce, F. A., Liverman, M. R., & Gay, A. (1991). Teachers' ability to interpret standardized test scores. *Educational Measurement: Issues and Practice, 10* (4), 16-18.

Impara, J. C., & Plake, B. S. (In press). Comparing counselors', school administrators', and teachers' knowledge in student assessment. *Measurement and Evaluation in Counseling and Development.*

Infantino, R. L. (1976). *Testing and accountability: A survey of the knowledge and attitudes of New York State secondary school English teachers.* Doctoral dissertation, State University of New York at Buffalo. (University Microfilms No. 77-614).

James C. Impara is Professional Associate, Buros Institute of Mental Measurements, University of Nebraska-Lincoln, Lincoln, NE.

ERIC Digest # EDO-CG-95-1

Assessment Skills for School Counselors

William D. Schafer

Perhaps the most controversial area within counselor education is that of assessment. Following Shertzer and Linden (1979), assessment is used here to mean methods or procedures that are employed to obtain information that describes human behavior. The purpose of this digest is to describe school counselors' roles in the area of assessment. Following a historical review of testing in counseling, some findings of a study by Schafer and Mufson (1993) that described roles employers require school counselors to perform are discussed. Conclusions are related to improving quantitative literacy in counselor education.

Historical Perspective

Knowledge needed by counselors to obtain evidence, evaluate its usefulness, and interpret its meaning have long been and continue to be debated. According to Minor and Minor (1981), that debate arose, in part, from the adoption of a humanistic perspective by many counselors and counselor educators, leading to a de-emphasis of models of counseling that entail quantitative assessment. In the 1960s, tests were viewed positively and were used primarily to identify students of outstanding abilities (Zytowski, 1982). However, in the early 1970s, Goldman (1972) suggested, using a well-known metaphor, that the marriage between tests and counseling had failed. At about that time, courts prohibited some established tests for certain purposes and legislatures passed bills to regulate aspects of the use of standardized tests. The validity and practical utility of all testing and appraisal techniques were questioned and negative consequences of "labeling" were emphasized.

Yet assessment remained commonplace in schools. Consider these findings in a survey by Engen, Lamb, and Prediger (1981) and reported by Zytowski (1982): 93% of secondary schools administered at least one test to all students; 76% administered achievement test batteries; 66% administered academic aptitude or intelligence tests; and 16% administered inventories of school or social adjustment or personality tests. By the 1980s, vocational guidance, according to Zytowski (1982), had become a unifying force between counseling and testing.

Zytowski (1982) described several changes that had been made in tests themselves and in their uses in counseling. One of these was an erosion of reliance on predictive validity and an accompanying emphasis on convergent and discriminant validity, along with construct validity. He also described the value of an assessment in terms of its ability to guide and motivate a professional toward seeking additional information for decision making. De-formalizing assessment, another change, included increased use of one-item measures, informed self-estimates, and card sorts or inventories in which quantified outcomes are less important than is the process the client engages in. Computers had become more instrumental in testing, moving from primarily scoring and score reporting to actual test administration and providing immediate feedback. Availability and interest in computer testing have clearly increased in the decade since Zytowski's summary appeared.

The counseling community has become more aware of ethical issues in testing. An American Counseling Association (ACA) statement titled Responsibilities of Users of Standardized Tests (RUST), published in 1978 and revised in 1989, urges awareness of differing purposes for testing and reminds us to consider the limitations of tests for any purpose and to evaluate the costs of not testing or using alternative methods of gathering the information needed.

Job Descriptions of School Counselors

In their study of skills needed by school counselors, Schafer and Mufson (1993) reviewed job analyses conducted by five school districts in five different states. They found a natural division of the job role expectations of school counselors into six areas: counseling (individual and group), pupil assessment, consultation, information officer, school program facilitator, and research and evaluation. There are assessment-intensive aspects of each of these.

The counselor's major function in the school is to counsel students individually and whenever practical in small or large groups.The counselor also is responsible for identifying students with special needs. These activities include interpreting test scores and non-test data.

Pupil assessment includes scheduling and preparing for testing, scoring tests or sending them out for scoring, recording results, and scheduling for interpretation. Counselors are also responsible for assisting students in evaluating their aptitudes and abilities through interpreting standardized tests. They may be expected to advise teachers who need to understand psychological evaluations and who are interested in improving their content-referenced testing skills.

The third function is that of a consultant. The counselor consults with and advises teachers, parents, and administrators in guidance matters and test score interpretation. In some schools the counselor helps teachers with psychological evaluations and content-referenced testing and advises school committees in selection of tests.

The function of information officer includes informing parents, teachers, and staff about counseling services, informing employers and colleges about students according to school policy, and ensuring two-way communication between school and home. Many of these activities involve test interpretation.

The fifth function is administrative, including school administration and counseling administration. Within school administration, the counselor is responsible for administering tests. Within counseling administrative functions, the counselor is expected to analyze guidance services. Also, the counselor is often asked to participate in decisions about the instructional curriculum.

The sixth function is research and evaluation. The counselor may be responsible for evaluating the school guidance program. The counselor is also expected to read and interpret literature, to apply research findings to everyday counselees' situations, and to improve his or her skills continuously through evaluation of counseling techniques.

The counselor responsibilities identified by Schafer and Mufson (1993) would likely be found in the large majority of school districts across the nation. Within the area of assessment, roles include test interpreter, test developer, evaluator of programs, consultant, and researcher. Several studies reviewed by Schafer and Mufson (1993) were supportive of these roles.

Assessment Skills Required by School Counselor Roles

The roles that have been identified imply that counselors should have certain skills related to assessment. Schafer and Mufson (1993) organized these into three areas: doing pupil assessment, doing program evaluation, and using basic research.

Doing pupil assessment encompasses: types of assessment; assessment systems and programs; test administration and scoring; test reporting and interpretation; test evaluation and selection; design, analysis, and improvement in instrument development; formal and informal methods of assessment; methods for using assessment in counseling; administrative uses of assessment; computer-based applications; and ethics of using assessments.

Doing program evaluation includes needs assessment; formative and summative evaluation; sources of evaluation research invalidity (instrumental, internal, and external); choosing evaluation designs; choices of and computational methods for descriptive and inferential statistics; writing evaluation proposals and reports; disseminating information; and research ethics.

Using basic research includes locating and obtaining relevant research reports; reading and summarizing research reports; evaluating validity of instruments and research designs; and understanding the purpose and assumptions of common inferential statistical procedures.

Conclusions

Schafer and Mufson (1993) generated aspects of school counselors' roles that are related to assessment. They also generated a list of assessment-related content areas in the CACREP standards that pertain to school counselor education programs. In order to study the fit of these two lists, for each job-definition role, they reviewed those CACREP content areas that seemed supportive of it. They concluded that these CACREP skills, conscientiously presented in a counselor education program,would in most areas constitute an adequate preparation for a beginning-level school counselor.

Focusing on the role of test interpreter, however, Goldman (1982) found little research evidence that tests as they have been used by counselors have made much of a difference to the people they serve. He felt the reasons for this are that counselors have not been prepared adequately to understand psychometric evidence, and that the predictive validity of test information is inadequate to support individual interpretation. He suggested that schools and other institutions should reduce the use of standardized tests and replace them with less formal and less quantitative methods. However, the implications for assessment in counselor education programs of such a shift are unclear. It seems unlikely that formal assessment methods will disappear from schools.

Perhaps, as Daniels and Altekruse (1982) observed, lack of integration of assessment and counseling rests on counselor educators' failure to provide integrating guidelines in both assessment and counseling coursework. Among other recommendations, they concluded that counselor educators should become more responsible for teaching assessment content as well as for demonstrating its interrelations with counseling in their other courses. Shertzer and Linden (1982) have suggested that a more systematic approach to counselor education at both the preservice and the inservice levels can produce professionals who are more sophisticated in the practice of assessment and appraisal. The same seems true in the areas of program evaluation and basic research.

References

Engen, H. B., Lamb, R. R., & Prediger, D. J. (1981, April). *Are secondary schools still using standardized tests?* Paper presented at the American Personnel and Guidance Association Convention.

Goldman, L. (1972). Tests and counseling: The marriage that failed. *Measurement and Evaluation in Guidance, 4*, 213-220.

Goldman, L. (1982). Assessment in counseling: A better way. *Measurement and Evaluation in Guidance, 15*, 70-73.

Minor, B. J. & Minor, J. H. (1981). A theoretical model for humanistic counseling research. *Personnel and Guidance Journal, 59*, 502-506.

Schafer, W. D. & Mufson, D. (1993, March). *Assessment literacy for school counselors.* Paper presented at the American Counseling Association Convention. (ERIC Document Reproduction Service No. TM 019 926)

Shertzer, B. & Linden, J. (1979). *Fundamentals of individual appraisal: Assessment techniques for counselors.* Boston: Houghton Mifflin.

Shertzer, B & Linden, J.(1982). Persistent issues in counselor assessment and appraisal. *Measurement and Evaluation in Guidance, 15*, 9-14.

Zytowski, D. G. (1982). Assessment in the counseling process for the 1980s. *Measurement and Evaluation in Guidance, 15*, 15-21.

William D. Schafer is Associate Professor in the Department of Measurement, Statistics, and Evaluation at the University of Maryland, College Park.

ERIC Digest # EDO-CG-95-2

Mental Health Counseling Assessment: Broadening One's Understanding of the Client and the Clients Presenting Concerns

Gerald A. Juhnke

Assessment has experienced a resurgence in recent years, both in the United States and abroad (Piotrowski & Keller, 1992; Watkins, 1994). Some continue to use the terms *assessment* and *testing* interchangeably. Both are vitally important to the counseling process (Lambert, Ogles, & Masters, 1992). Yet, assessment is broader in scope than testing. Typically, assessment includes gathering and integrating information about a client in a manner that promotes effective treatment (Cohen, Swerdlik, & Smith, 1992). This can be accomplished by using testing in conjunction with other methods, such as qualitative techniques, behavioral assessments, and review of past client records. Testing should not be used as the only source of information about a client (Anastasi, 1992).

Corroborating data from a number of sources helps create a more thorough understanding of the client and his or her presenting concerns. The counselor can then interpret these data and formulate hypotheses related to the client's strengths and weaknesses. Data gathered and the hypotheses formed thereby contribute to the creation of an effective counseling strategy. This digest discusses how counselors can use assessment as a continuous process throughout treatment. It also reviews three common forms of assessment which can be used in conjunction with testing.

Continuous Assessment

Vacc (1982) notes, "Assessment in counseling should be viewed not as a one-time prediction activity but rather as continuous throughout the counseling process..." (p. 40). Continuous assessment influences the direction of treatment in two ways. First, presenting concerns and client circumstances are not static. Goals identified by the client during the initial assessment often must be modified or reordered to meet new and urgent client needs. Continuous assessment apprises the counselor of possible new and urgent needs which have arisen since the initial assessment. These needs can then be addressed through the counseling process. Second, assessment can aid in evaluating the efficacy of treatment. Upon entering treatment, an initial assessment establishes the client's baseline of functioning. Continuous assessment allows comparisons between this initial baseline and the client's current functioning. Improvements suggest treatment efficacy and the benefit of continuing the current treatment course. Reduction in functioning or a lack of improvement, however, suggests a need to alter treatment. Continuous assessment, therefore, is important, because it keeps the counselor apprised of the client's ever-changing needs and indicates treatment efficacy.

Qualitative Assessment

Qualitative assessment techniques are compatible with the belief that "...assessment activities should not stand outside the change process; rather, they should blend into treatment strategies to guide self-discovery and to inform clients" (Drum, 1992, p. 622). Unlike standardized tests, qualitative assessments often consist of games or simulation exercises that are flexible, open-ended, holistic, and nonstatistical (Goldman, 1992). Typically a debriefing follows the qualitative assessment experience. Clients can process what they learned from the experience immediately within the counseling session.

One commonly used qualitative assessment experience is called, "The Life Line" (Goldman, 1992). The intent of this experience is to help clients reflect upon significant past events which have influenced them. Clients draw a horizontal timeline on a blank sheet of paper. They are then asked to recall past significant experiences, relationships, events, or wishes which have influenced their lives, and to plot these along the timeline. The result gives the counselor detailed information about significant events in the client's developmental history.

Similarly, role plays can serve as a qualitative assessment experience. For example, a mental health counselor may ask a client to role play a recent anxiety-provoking experience (e.g., an argument with a supervisor, receiving a speeding ticket, etc.). The role play provides the mental health counselor with a sample of the client's behaviors. As the role play is being demonstrated the counselor can query the client regarding possible negative self-talk (e.g., I'm so stupid, he'll never listen to me, etc.). Understanding the self-talk used by a client can help the counselor generate effective intervention ideas. Clients can also practice new counselor-directed behaviors or self-talk (e.g., I'm intelligent, he'll want to listen to me) within the counseling session through role plays.

Another qualitative assessment technique that can provide valuable information is a photograph safari. Depending upon the presenting concerns, the counselor may request that the client bring to the session photographs of the client's family-of-origin or childhood. The counselor and client can jointly review these photographs. Particular attention should be paid to: (a) those present in the photographs; (b) those consistently absent from the photographs (e.g., Are the client's siblings always included in the photographs but the client absent?); (c) common themes of the photographs (e.g., Are all the pictures taken on the family farm? Are pictures only taken during certain holidays?); (d) proximity to significant others posing in the photographs (e.g., Is the client consistently posed beside the client's father? Is the client consistently standing apart from other family members?); and (e) emotions displayed on family members' faces (e.g., Does the client consistently pout or appear angry in photographs?). Such qualitative assessment techniques can promote insight for the client and therapeutic direction for the counselor.

Behavioral Assessment

Counselors using behavioral assessments are most interested in recording manifest behaviors. Emphasis is placed upon identifying antecedents to problem behaviors and consequences that reduce their frequency or eliminate them (Galassi & Perot, 1992). Both indirect and direct methods are used for behavioral assessments. Indirect methods of behavioral assessment might include the counselor interviewing the client or talking to significant others about the reported problem behavior. Indirect behavioral assessment provides important information about the client and the client's presenting concerns, but the information obtained may be contaminated by misperceptions or biases about the client or the client's behaviors. More direct methods reduce the probability of misperceptions or biases, and might include counselor observation of the client or client self-monitoring. A behavioral problem checklist or procedures especially designed to record the client's concerns directly (e.g., recording the frequency, duration, and intensity of marital arguments) can be used to help clarify possible antecedents to behavioral problems and record what subsequent interactions result in their discontinuance.

Past Records

Reviewing previous client records (e.g., counseling, school, police, medical, military, etc.) can help the mental health counselor identify important patterns which the client may be unaware of or disinclined to discuss readily (e.g., problems with authority figures, self-injurious behaviors occurring after the ending of significant relationships, etc.). These records can be a vital source of information. Often a review of previous counseling records will indicate what types of treatment were attempted. Previously ineffective treatments can be ruled out, and treatment regimes found helpful re-implemented.

Concomitantly, past records link the client's history to the presenting concern. A counselor can gain increased clarity about the immediate concern based upon an improved understanding of previous stressors or transitions leading to the client's current condition. The counselor can then address the cause(s) of the symptoms rather than the symptoms themselves.

Summary

Assessment provides direction for treatment and aids in the evaluation process. Although many methods can be employed to promote a thorough assessment, no one method should be used by itself. Ultimately, it is the counselor's responsibility to gain sufficient information regarding the client and the client's presenting concerns to establish an effective treatment strategy. Using a combination of assessment techniques increases the likelihood of positive interventions and promotes successful treatment.

References

Anastasi, A. (1992). What counselors should know about the use and interpretation of psychological tests. *Journal of Counseling & Development, 70*(5), 610-615.

Cohen, R. J., Swerdlik, M. E., & Smith, D. K. (1992). *Psychological testing and assessment: An introduction to tests and measurements.* Mountain View, CA: Mayfield Publishing Company.

Drum, D. J. (1992). A review of Leo Goldman's article "Qualitative assessment: An approach for Counselors". *Journal of Counseling & Development, 70*(5), 622-623.

Galassi, J. P. & Perot, A. R. (1992). What you should know about behavioral assessment. *Journal of Counseling & Development, 70*(5), 624-631.

Goldman, L. (1992). Qualitative assessment: An approach for counselors. *Journal of Counseling & Development, 70*(5), 616-621.

Lambert, M. J., Ogles, B. M., & Masters, K. S. (1992). Choosing outcome assessment devices: An organizational and conceptual scheme. *Journal of Counseling & Development, 70*(4), 527-532.

Piotrowski, C., & Keller, J. W. (1992). *Projective techniques: An international perspective*. (ERIC Document Reproduction Service No. ED 355 273).

Vacc, N. A. (1982). A conceptual framework for continuous assessment of clients. *Measurement and Evaluation in Guidance, 15* (1), 40-47.

Watkins, C. E., (1994). Thinking about "Tests and Assessment" and the career beliefs inventory. *Journal of Counseling & Development, 72*(4), 421-423.

Gerald A. Juhnke is a Professor and Clinic Coordinator in the Department of Counseling and Educational Development in the School of Education at the University of North Carolina at Greensboro.

ERIC Digest # EDO-CG-95-3

CACREP Accreditation: Assessment and Evaluation in the Standards and Process

Carol L. Bobby & Joseph R. Kandor

The Council for Accreditation of Counseling and Related Educational Programs (CACREP) was organized in 1981 as the accrediting agency responsible for reviewing and evaluating counseling and student affairs practice in higher education programs against a set of nationally recognized standards. Its incorporation as an independent body was the culmination of years of work by the American Counseling Association (ACA) and its divisions to define the knowledge and skills required for entry into the profession and to advocate that these requirements be implemented by the preparation programs offering counseling and student affairs practice degrees. Striving to foster excellence in its organizational development, CACREP sought an external review and evaluation of its own accrediting practices by the Council on Postsecondary Accreditation (COPA) and was awarded recognition status by this body in 1987. This recognition has been maintained and was recently transferred to the Commission on Recognition of Postsecondary Accreditation (CORPA) in a recent restructuring of the recognition function.

As a CORPA-recognized accrediting agency, the CACREP accreditation process must incorporate a pattern of review that includes integral self-study of the program against nationally accepted criteria, followed by an on-site visit by an evaluation team, and a subsequent review and accreditation decision rendered by a central governing group. This pattern illustrates the important role that assessment and evaluation play in every accreditation agency's process. The purpose of this digest is to explore the specific levels of assessment and evaluation involved in the CACREP accreditation process, as well as to provide an overview of curricular experiences in assessment and evaluation.

Assessment and Evaluation in the CACREP Review Process

The CACREP review process is a multilevel assessment and evaluation process with four basic levels of review occurring simultaneously. These levels are: (l) the program's internal assessment and evaluation of how the CACREP standards are implemented; (2)

an external review of the program by CACREP to determine compliance with the standards; (3) regular and systematic program evaluation based upon the program's own mission and objectives; and (4) regular and systematic evaluation of CACREP's accreditation process based upon its mission and objectives. As a voluntary activity, a program's participation in the CACREP review process speaks to a high level of commitment to quality assurance in program delivery to students.

An Overview of the Accreditation Review Process

The first two levels listed are integral to the CACREP accreditation process, as they involve the three general requirements of self-study, on-site evaluation, and final decision review that are common to all accrediting agency reviews. Within these three general requirements, however, are many points at which assessment and evaluation occur. During the internal review, a program begins to examine itself against the CACREP eligibility requirements and standards. Following a preliminary review of the accreditation standards, a program will need to assess resources such as faculty support, institutional support, administrative support, budgetary support, clinical facilities, library and research facilities and services, and student enrollments to determine if the benefits of pursuing accreditation at this time can be balanced with the costs and potential need to reallocate resources. If the answer is yes, the program begins a more formal internal evaluation of how each of the CACREP standards is perceived to be met. The results of this evaluation are compiled in a self-study report, appended with supporting documentation, and submitted to CACREP for an external review.

The external review also requires several steps. First is an initial review of the self-study report by a subcommittee of CACREP's governing board. If the reviewers concur that the self-study adequately addresses the eligibility requirements and standards, a site visit is recommended. The next phase of review involves sending a team of trained volunteers to the campus to validate the responses provided in the self-study report. The team spends a minimum of three days at the program's campus evaluating the program against the standards by reviewing additional documentation, visiting relevant facilities and sites, and interviewing students, graduates, faculty, administrators, and clinical supervisors. At the completion of the on-site visit, the team members submit a report to CACREP that reflects their evaluation of the program's compliance with the standards. Programs are then allowed to review the team's report and respond to the relative accuracy of its content.

The final step of the external review entails an evaluation of a program's compliance with the standards by CACREP's governing body, the Board of Directors. At this point in the process, the board has access to the original self-study report, the team's report, the program's response to the team's report, and all appended documentation. A thorough review of all documents occurs and a decision is rendered to award or deny accreditation to the program.

Program Evaluation Requirements

In addition to the internal and external evaluation that occurs as part of the CACREP accreditation review process, a program seeking accreditation must document its own process of regular and systematic program evaluation. According to Section VI: Evaluation in the Program of the CACREP Standards, such evaluations should include:

1. developmental, systematic assessment of each student's progress throughout the program with consideration given to academic performance, professional development, and personal development;
2. internal and ongoing reviews by program faculty of curricular offerings, objectives, professional trends, student learning outcomes, and types of students seeking admission to the program;
3. external review through follow-up studies with graduates of the program to assess their perceptions and evaluations of major aspects of the program;
4. assessment of perceptions about the program among employers of graduates, field placement supervisors, and cooperating agency personnel, and;
5. assessment of faculty by currently enrolled students.

These criteria insure that evaluation become an important component of every CACREP-accredited program and that the evaluation involves multiple levels of assessment.

Evaluation of CACREP

The fourth level of evaluation important to the CACREP accreditation process involves evaluating the evaluators; that is, CACREP's examination of itself through internal and external review mechanisms. Just as programs seeking accreditation are required to complete a self-study report that addresses a program's compliance with standards, CACREP must periodically review itself against a set of criteria that represent good accrediting practices. Documenting compliance with these criteria is the basis for receiving and maintaining status as a recognized accrediting agency. Similar to the CACREP process, the recognition process entails a series of review steps before final decisions are rendered.

In addition to this review process, CACREP has also put into place a series of procedures to allow for regular and systematic review of its activities. A complete review of the CACREP standards must occur every seven years. It is at this time that CACREP seeks comment from educators, practitioners, administrators, supervisors, students, graduates, employers, and the public regarding the content of its curricular, clinical, institutional, and program requirements. Rationales for recommended changes with an assessment of the impact of the recommended changes are reviewed by a committee, and drafts with recommended changes circulated for further comment. The process entails a series of draft review and comment periods prior to final adoption of revised standards. The process ensures that the standards remain responsive to the needs of a dynamic and changing society.

Other avenues of evaluation include assessing a program's satisfaction with the conduct of the on-site team visit, as well as the program's perceptions of both the site visitors' and CACREP staff's understanding of the philosophy of accreditation and knowledge of the standards. CACREP staff have also conducted periodic research on issues such as inhibitors to seeking accreditation and the frequency with which certain standards are cited for noncompliance in the final accreditation review.

Curricular Experiences in Assessment and Evaluation

Within a CACREP program, students must complete coursework related to assessment and evaluation. Standard II.J.6: Appraisal and Standard II.J.7: Research and Program Evaluation in the CACREP Accreditation Standards Manual (1994) outline the specific curricular experiences required of every student in the program.

Appraisal includes studies that provide an understanding of individual and group approaches to assessment and evaluation. Research and program evaluation requires studies that provide an understanding of types of research methods, basic statistics and ethical and legal considerations in research.

Within doctoral programs, students must receive curricular experiences that represent an extension of the requirements for appraisal and evaluation outlined above (Doctoral Standard II.A). The program must also provide the doctoral student with advanced preparation in design and implementation of quantitative and qualitative research and methodology (Doctoral Standard II.C.4) and models and methods of appraisal (Doctoral Standard II.C.5).

Summary

The concepts of assessment and evaluation are significant throughout the CACREP accreditation process and standards. Assessment and evaluation remain at the heart of the practice and procedures used by CACREP to ascertain whether a program will achieve or be denied accredited status. Furthermore, the accreditation standards themselves require that not only are students provided knowledge and skills in the areas of assessment and evaluation, but that programs regularly and systematically assess these types of curricular offerings, along with other aspects of program operations.

The importance of CACREP undergoing periodic evaluation of its own accrediting practices and standards should not be underestimated. It is the combination of results from each of the various levels of evaluation described in this paper that allow for quality assurance in program development while simultaneously embracing change so that entering professionals will have the knowledge and skills necessary to deal with our rapidly changing society

References

Council for Accreditation of Counseling and Related Educational Programs. (1994). *Accreditation procedures manual and application*. Alexandria, VA: Author.

Carol L. Bobby is Executive Director of CACREP, Alexandria, Virginia.

Joseph R. Kandor is Chair and Professor in the Department of Counselor Education at State University of New York at Brockport.

ERIC Digest # EDO-CG-95-4

The Role of Assessment in Counselor Certification

Thomas Clawson

Certification of professional counselors is presently viewed in two realms, those of state regulation and of national voluntary credentialing. Many states use the term *certification* in two contexts, school counselor certification and certification to practice counseling privately for a fee. In this digest, we will consider national voluntary certification only.

The first national certification began in 1972 with the incorporation of the Commission for Certification of Rehabilitation Counselors. In 1979, the National Academy for Certified Clinical Mental Health Counselors began certifying counselors trained in the specialty of clinical mental health counseling. Soon thereafter, in 1984, the National Vocational Guidance Association (now the National Career Development Association) began certifying career counselors. In 1983, the National Board for Certified Counselors (NBCC) began certification for general practice counselors. And, as this digest is being written, the International Association of Marriage and Family Counselors is beginning a certification process. Clinical mental health counselors and career counselors have merged with the National Board for Certified Counselors to become a specialty certification of the general practice of counseling.

Across the realm of certifications in the counseling profession is the common thread of assessing individual counselors, training, supervision, experience, and knowledge; the similarities across the processes are remarkable.

Methods of Assessment

Counselor certification begins with individuals providing certification boards with a portfolio of data pertaining to their training, supervision, experience, and knowledge. All are areas of difficulty in quantifying or qualifying.

Training

Training is perhaps the easiest certification area to assess but even in evaluation of coursework, a variety of factors are evident. Most academic training reviews require determination of term (semester, trimester, quarter) hours awarded for graduate study in regionally accredited institutions. Course titles of counseling and related disciplines number in the thousands. Certification boards must categorize courses by reviewing catalogue course descriptions or syllabi. While quantifying transcript review appears to be a simple task, it consumes a great proportion of portfolio review time.

A further complication in determining appropriate training appears when certifying boards accept nontra-ditional education. Processes must be developed that compare home study and other methods of delivery with traditional campus experiences. This may be done by designating which areas of study must be delivered by traditional professor/student/classroom methods and which courses may safely use nontraditional techniques such as distance learning. In counseling, the most important training dynamic is the demonstration of theory-to-practice transference. Topics requiring application of skills to counselees, such as group, individual, or family counseling and assessment of individuals or groups indicate the need for close supervision by a professor.

Supervision

Supervision duration is easily assessed if certification boards can define supervision and supervisors clearly. Then accurate reporting of supervision by supervisors establishes an hour total to judge against a standard number of hours. As the concept of certification has matured, the qualification and definition of supervision have advanced. Defining and assessing supervision, however, is probably the least sophisticated and standardized certification area assessed at present. Bernard and Goodyear (1992) point out that as models of supervision grow, the research and practice will bring forth clearer definitions.

Experience

Experience is easily quantified for assessment once standards and permutations are set. For example, certification boards may set a year or hour experience requirement and also set ways to accumulate hours of supervised experience at less than full-time employment. Again, as certification evolves the ways of achieving experience have become more strict. In counseling, this is probably a result of the maturation of the profession.

Knowledge

Knowledge is relatively simple to assess if the universe of the information to be assessed is small. Counseling information included in the eight core areas of the Council for Accreditation of Counseling and Related Educational Programs are as follows: (1) Human growth and development; (2) Social/cultural and family foundations; (3) The helping relationship (including counseling theories); (4) Group dynamics, processes, and counseling; (5) Lifestyle and career development; (6) Appraisal of individuals; (7) Research and evaluation; and (8) Professional orientation. These core areas are an example of the discipline producing more and more information as the research and literature base of counseling grows. Therefore, sampling the relevant knowledge base becomes an increasingly difficult task. All counselor certification examinations employ multiple-choice, single-answer formats and range from 100 to 250 items per form.

Because the practice of counseling involves application of information to action, examination constructors face the task of applying knowledge data to cases or situations. The standard beginning point for this application is the job analysis or study of behaviors used in a profession. Most counselor certification exams are based upon comprehensive job analyses of practicing counselors. The National Organization for Competency Assurance requires state-of-the-art job analyses as a prerequisite for accreditation of certification programs (National Organization for Competency Assurance, 1993). Professional examinations which are not based upon comprehensive study of the necessary behaviors needed for professional practice are suspect even before reliability and validity statistics are gathered.

Job Analysis

Shimberg and Rosenfield (1990) identify the general purpose of job analyses as: a process that seeks information from a large number of incumbent practitioners regarding the most important aspects of the job; and the knowledge and skills needed to perform the job in a safe and effective manner (p. 14).

Fine (1986) continues that job analyses can also provide definition of the behaviors needed to practice, knowledge and abilities needed in training curricula, and relevant assessments of performance (p. 55).

Loesch and Vacc (1991) describe job analyses as having multiple facets to obtain a picture of a profession. Three major categories of decisions must be considered in conducting a job analysis: (1) conceptual; (2) procedural; and (3) analytical. Conceptual decisions as a basis for a credentialing examination are intended to allow for development of a "test blueprint." Procedural decisions include research methodology, type of examination format, and item generation technique. Analytical decisions involve the statistical and methodological treatment of the list of professional behaviors generated (pp. 5-6).

So, job analysis is not directly applied to the individual applicant for certification, but to a large group of practicing professionals. It is the precursor to assessment of certificants and, indeed, essential for logical application of certification criteria.

Continuing Training

Continuing training is an ongoing assessment process that begins, for certification purposes, after credentialing is achieved. Most certifying boards require continuing education as a part of recertification. Some require both continuing education and re-examination periodically. The NBCC requires twenty clock hours of continuing education per year over each five year certification period. All certificants must attest to continuing their training and submit to random inspection.

Recommendations

Every national program certifying counselors uses multiple-choice examinations as part of the application requirement. While this method can assess information retention readily, it does not lend itself to measuring counseling skills and application of theory to skills. Recent revisions of the National Counselor Examination for Licensure and Certification (NCE) have included more applied items. Future modifications should include methodologies that assess skills better. Tape simulations, computer applications, branching answer format, in vivo review, and case scenario models all may be included in future revision. These modifications, of course, have expense implications, which has been the major force in retention of multiple-choice formats in counselor certification.

In an emerging profession such as counseling, an examination which is not undergoing change will soon be obsolete. Monitoring of professional practice, research, and literature, as well as advances in examination development and theory are essential to a good assessment program.

The Clinical Mental Health Counselor Academy of the NBCC has always required a tape sample of counseling with a current counselee. This method requires extraordinary time expenditure by applicants for certification as well as tape reviewers. Each tape is reviewed by clinical counselors to assure clinical counseling skills. Clearly this process demands the most scrutiny of reliability (interrater in this case) of all NBCC processes. Ongoing reliability checks of tape review processes are a must. More research will no doubt help delineate better methods of judging tape samples.

Since NBCC has been gathering data on counselor behavior and examination statistics for over twelve years, the time has come to begin releasing these assessment data for use by those with interest in the profession. Such a process is now occurring beginning with the release of all data regarding the most recent and comprehensive job analysis performed within the counseling profession.

Requiring supervision for certification continues to generate a need for better definitions of supervision and qualification of supervisors. In a profession depending upon performance, supervision of pre-service and in-service counseling is essential. Not only will standards need to be developed further, but some more quantifiable measures of supervision must emerge.

Summary

While counseling is an emerging profession, the NBCC has kept pace with national mandates for state-of-the-art assessment techniques. Present methods are constantly being modified in light of assessment advancements. Use of presently unreported data may lead to further positive steps in selecting certificants.

References

Bernard, J. M. & Goodyear, R. K. (1992). *Fundamentals of clinical supervision*. Needham Heights, MA: Allyn and Bacon.

Fine, S. A. (1986). Job analysis In R. A. Berk (Ed.), *Performance assessment: Methods & Applications* (pp. 53-81). Baltimore, MD: Johns Hopkins University Press.

Loesch, L. C. & Vacc, N. A. (1991). *National counselor examination technical manual (1991 rev.)*. Greensboro, NC: National Board for Certified Counselors.

Shimberg, B. & Rosenfeld, M. (1990, Winter). Psychometric issues in job analysis: Key to content valid tests. *CLEAR Exam Review,* pp. 14-15.

Thomas Clawson, Ed.D, NCC, is the Executive Director of the National Board for Certified Counselors in Greensboro, North Carolina.

ERIC Digest # EDO-CG-95-5

Assessment of Counselor Performance

Larry C. Loesch

Assessment of counselor performance is directly linked to assessment of counseling outcome because, presumably, counseling outcome is contingent upon counselor performance. Thus, the assessment of counseling outcome literature is the general context for the more specific literature on assessment of counselor performance, and the same major themes are evident in both arenas. Historically, counselor performance has been assessed, either directly or *vis-a-vis* outcome, primarily in regard to actual *counseling* service rendered through assessments by counselors themselves, their clients, or external evaluators. However, recently, non-counseling activities also have been assessed as part of the overall evaluation of counselor performance.

Many methodologies have been used to assess counselor performance, including assessments such as interviews, linguistic content analyses, simulations, self-reports, applications of behavioral criteria, and rating scales. The focus of these assessments has ranged from the global to the specific. Rating scales are the most commonly used method, but no assessment procedure has emerged as most psychometrically appropriate, reliable, valid, or effective.

Counselors' Self-Assessments

The (Rogerian) premise that effective counseling necessitates substantial emotional congruence between counselor and client is widely espoused in the counseling profession. The highly personal nature of such emotional congruence suggests that the counselor is the best person to assess it. Thus, a variety of methods, such as "learning diaries," self-rating scales, or audiotaped "introspective dialogues," have been used to allow counselors to indicate the degree to which they have achieved emotional congruence with their clients.

Counselor self-assessments are popular among counselors, and arguably valuable, for purposes of self development and improvement. However, because of their subjectivity, their results rarely have been generalizable. Also, the methodologies generally have not withstood psychometric scrutiny. Therefore, counselor self-assessments are not widely used for effective assessment of counselor performance.

Assessments by Clients

Because counseling is *for* the client, it is a reasonable assertion that the client is the person best able to assess the degree to which the counselor has performed effectively. The credence of this assertion is evident in that client assessment of counselor performance is widely used, and many methodologies have been developed to facilitate it. In general, clients have been requested to assess counselor performance in regard to the counselor being or behaving in a helpful way or the degree of the client's personal change.

A counselor's "helpfulness" has been most frequently assessed by clients through use of post-counseling "debriefing" interviews or rating scales. Typically assessed is the client's perceptions of the counselor's personal dynamics (e.g., degree of caring) or actions or behaviors which were helpful. The focus has often been on the latter, but some suggest it should be on the former (Herman, 1993).

Some rating scales have been developed to allow clients to assess counselors' personal dynamics. However, most are intended to allow client evaluation of the extent to which the counselor engaged in behaviors (particularly verbalizations) presumed or established to be related to counseling effectiveness. Some of these instruments have been shown to have quite good psychometric properties. Quality issues aside, however, use of rating scales completed by clients is one of the two most common methods of assessment of counselor performance.

Client self-assessment of change as an indicator of counselor performance typically has involved commentary, ratings, or self or other reported behavior changes. Unfortunately, however, these procedures have been used only infrequently for assessment of counselor performance, probably because the best data are obtained some time after counseling has been terminated.

Assessments by External Evaluators

Assessment of counselor performance by persons external to the counseling relationship is by far the most frequently used approach. The obvious advantage of such assessments is greater objectivity. In addition, external assessments usually are psychologically and behaviorally less intrusive, particularly if the assessments are applied to audio or video tape-recorded counseling. External assessments also may be more practical because they are more easily applied to different types of counseling (e.g., individual, group, or family) or specific counseling contexts (e.g., see Ponterotto, Rieger, Barrett, & Sparks, 1994).

A wide variety of external assessment methodologies have been employed, including some only infrequently used in the counseling profession such as content analyses, critical incident techniques, or computer simulations (McLeod, 1992). However, rating scales again are the most frequently used assessment method. Rating scales have been developed to assess many different aspects of counselor performance, but most are focused upon the frequency and/or effectiveness of counselors' use of specific and behaviorally defined counseling skills.

The results of external assessments of counselor performance have been used in the context of both formative and summative evaluations. In the formative context, rating scales completed by counselors' supervisors, peers-in-training, or professional colleagues are often used on some regularly scheduled basis to provide process or skill development feedback to the counselors assessed. In the summative context, results from rating scales completed by supervisors, colleagues, or researchers are often used for program or personnel evaluation or research purposes.

Assessment of Non-Counseling Functions

The most recent trend in assessment of counselor performance has been to broaden the perspective on what it means to be an effective counselor, that is, to acknowledge that there is more to being a good counselor than just counseling skill (Bell, 1990). Assessments within this perspective encompass both actual counseling performance and other activities in which professional counselors engage. Assessments in the latter regard typically address activities such as diagnosis, case management, treatment planning, consultation, professional development, research, materials development, and interprofessional communications. These non-counseling components of counselor performance are typically assessed through use of rating scales by external evaluators. However, alternatives such as portfolio assessment or service recipient evaluations apparently are gaining favor.

Conclusion

It has long been recognized that good assessment involves multiple measurements of whatever is being assessed, and this principle has been recognized in regard to the assessment of counselor performance (Ridgway, 1990). There are literally hundreds of assessment instruments and techniques available to assess various facets of counselor performance. Therefore, it is not difficult to fulfill the multiple measurement criterion. Ironically, however, some experts have suggested that there are *too many* measures of counselor performance, a problem resulting from the many situation-specific assessment devices that have been developed. Most of these assessments are not derived from clearly defined constructs, are narrow in focus, and lack psychometric quality. Thus, comparability across measurements is restricted and generalizability across situations is limited.

The assessment of counselor performance will be enhanced when assessments are clearly and cogently described (Meier & Davis, 1990) and are used within an effective conceptual (evaluation) scheme (Lambert, Ogles, & Masters, 1992). Even more importantly, however, truly effective counselor performance assessment will be achieved when the assessments used fulfill accepted psychometric quality criteria (McLeod, 1992).

References

Bell, P. (1990). *Developing and implementing a counseling evaluation program.* (ERIC Document Reproduction Service No. ED 320633).

Herman, K.C. (1993). Reassessing predictors of therapist competence. *Journal of Counseling and Development*, *72*(1), 29-32.

Lambert, M.J., Ogles, B.M., & Masters, K.S. (1992). Choosing outcome assessment devices: An organizational and conceptual scheme. *Journal of Counseling and Development*, *70*(4), 527-532.

McLeod, J. (1992). What do we know about how best to assess counsellor competence? *Counselling Psychology Quarterly*, *5*(4), 359-372.

Meier, S.T., & Davis, S.R. (1990). Trends in reporting psychometric properties of scales used in counseling psychology research. *Journal of Counseling Psychology*, *37*(1), 113-115.

Ponterotto, J.G., Rieger, B.P., Barrett, A., & Sparks, R. (1994). Assessing multicultural counseling competence: A review of instrumentation. *Journal of Counseling and Development*, *72*(3), 316-322.

Ridgway, I.R. (1990). Multiple measures for prediction of counselor trainee effectiveness. *Canadian Journal of Counselling*, *24*(3), 165-177.

Larry C. Loesch, Ph.D., NCC, is a professor and graduate coordinator in the Department of Counselor Education at the University of Florida in Gainesville, Florida.

ERIC Digest # EDO-CG-95-6

Evaluating School Guidance Programs

Norman C. Gysbers

"Demonstrating accountability through the measured effectiveness of the delivery of the guidance program and the performance of the guidance staff helps ensure that students, parents, teachers, administrators, and the general public will continue to benefit from quality comprehensive guidance programs" (Gysbers & Henderson, 1994, p. 362). To achieve accountability, evaluation is needed concerning the nature, structure, organization, and implementation of school district/building guidance programs; the school counselors and other personnel who are implementing the programs; and the impact the programs are having on students, the schools where they learn, and the communities in which they live. Thus, the overall evaluation of school district/building guidance programs needs to be approached from three perspectives: program evaluation, personnel evaluation, and results evaluation (Gysbers & Henderson, 1994).

Guidance Program Evaluation

Guidance program evaluation asks two questions. First, is there a written guidance program in the school district? And second, is the written guidance program the actual implemented program in the buildings of the district? Discrepancies between the written program and the implemented program, if present, will come into sharp focus as the program evaluation process unfolds.

To conduct program evaluation, program standards are required. Program standards are acknowledged measures of comparison or the criteria used to make judgments about the adequacy of the nature and structure of the program as well as the degree to which the program is in place. For example, here is a program standard:

> The school district is able to demonstrate that all students are provided the opportunity to gain knowledge, skills, values, and attitudes that lead to a self-sufficient, socially responsible life. (Gysbers & Henderson, 1994, p. 481)

To make judgments about guidance programs using standards, evidence is needed concerning whether or not the standards are being met. In program evaluation such evidence is called documentation. Using the standard listed above, evidence that the standard is in place might include the following:

- a developmentally appropriate guidance curriculum that teaches all students the knowledge and skills they need to be self-sufficient and lead socially responsible lives.

- a yearly schedule that incorporates the classroom guidance plan (Gysbers & Henderson, 1994, p. 482).

Documentation of such evidence could include:

- guidance curriculum guides
- teachers' and counselors' unit and lesson plans
- yearly master calendar for the guidance program
- curriculum materials (Gysbers & Henderson, 1994, p. 482)

Sometimes the program evaluation process is called a program audit. The American School Counselor Association, for example, uses the term audit in its program evaluation materials. The Association has developed guidelines for a program audit for secondary schools (ASCA, 1986), for middle/junior high schools (ASCA, 1990b), and for elementary schools (ASCA, 1990a).

Guidance Program Personnel Evaluation

Personnel evaluation begins with the organizational structure and activities of the guidance program in a school district. A major first step is the development of job descriptions that are based directly on the structure and activities of a school district's guidance program.

Using the Missouri Comprehensive Guidance Program framework, for example, the job description of a school counselor would include the following key duties: implementing the guidance curriculum; counseling individuals and small groups concerning their educational and occupational plans; counseling individuals and small groups with immediate needs and specific problems; consulting with parents and teachers; referring students to appropriate community agencies; coordinating, conducting, and being involved with activities that improve the operation of the school; evaluating and updating the guidance program; and continuing professional development (Starr & Gysbers, 1993). (For examples of job descriptions of other guidance personnel including director of guidance, career guidance center technician, and high school registrar see Gysbers & Henderson, 1994, 422-428).

Guidance program personnel evaluation is based directly on their job task descriptions and usually has two parts: a formative part (supervision) and a summative part (evaluation). The job task description identifies the performance areas to be supervised and evaluated. Gysbers and Henderson (1994) have developed an extensive listing of job task descriptors for school counselors grouped under the basic guidance program components of guidance curriculum, individual

261

planning, responsive services, and system support plus the areas of professional relationships and professional responsibilities.

Program Results Evaluation

Having established that a guidance program is operating in a school district through program evaluation, and having established through personnel evaluation that school counselors and other guidance program personnel are carrying out the duties listed on their job descriptions 100% of the time, it now is possible to evaluate the results of the program. Johnson (1991) suggested that there are long-range, intermediate, immediate, and unplanned-for results that need consideration. According to Johnson, long-range results focus on how programs affect students after they have left school. Usually long-range results are gathered using follow-up studies. Intermediate results focus on the knowledge and skills all students may gain by graduation from participating in the guidance program. Immediate results are the knowledge and skills students may gain from participating in specific guidance activities. Finally, the possibility of unplanned-for results that may occur as a consequence of guidance activities conducted as a part of the guidance program also need to be taken into account.

For the purposes of this digest, illustrations of immediate and intermediate results evaluation using the structure of the Missouri Comprehensive Guidance Program model (Starr & Gysbers, 1993) are presented in the form of two research questions. First, do students master guidance competencies as a result of their participation in the Guidance Curriculum Component of the model (immediate evaluation)? Second, do students develop and use career plans as a result of their participation in the Individual Planning Component of the model (intermediate evaluation)?

Immediate Evaluation — Guidance Competency Mastery

Do students master guidance competencies? Johnson (1991) outlined the following procedures to answer this question for immediate results. First the competencies to be mastered need to be identified. Second *what* results (what students should be able to write, what they may be able to talk about, or what they may be able to do) are specified. Then *who* will conduct the evaluation is decided. This is followed by *when* the evaluation is done. Then *criteria* are established so that judgments can be made about students' mastery of guidance competencies. Finally, *how* all of this is done is specified.

Do students master guidance competencies? Another way to conduct immediate evaluation, to measure mastery of guidance competencies, is the use of a confidence survey. In this format, guidance competencies are listed and students are asked to rate how confident they are that they have mastered these competencies. The confidence survey can then be used as a pre-post measure. Gain scores can be obtained and related to such measures as academic achievement and vocational identity. (Gysbers, Hughey, Starr, & Lapan, 1992; Gysbers, Lapan, Multon, & Lukin, 1992; Lapan, Gysbers, Hughey, & Arni, 1993).

Intermediate Evaluation—Career Plans

Do students develop and use career plans? In making judgments concerning the career plans of students, criteria need to be identified as to what makes good plans. Four criteria are recommended; plans need to be comprehensive, developmental, student-centered and student-directed, and competency based.

Based on these criteria, one way to evaluate students' career plans is to judge the extent to which the activities included in the Individual Planning Component of the guidance program lead to the development of plans that meet these criteria. A second way is to make judgments about the adequacy of the plan contents. Finally, a third way is to judge their use. Do students actually use their career plans in planning for the future?

Summary

In order to fully evaluate comprehensive school guidance programs, three forms of evaluation are required. First, the program must be reviewed using program standards, evidence, and documentation to establish that there is a written guidance program in a school district and/or building and that the written program is the implemented program. Second, guidance program personnel need job descriptions derived directly from the program so that evaluation forms can be developed and used for formative and summative personnel evaluation. Third, results evaluation that focuses on the impact of the guidance and counseling activities in the guidance curriculum, individual planning, responsive services, and system support components of a comprehensive guidance program is mandatory.

References

American School Counselor Association (1986). *Professional development guidelines for secondary school counselors: A self audit.* Alexandria, VA: Author

American School Counselor Association (1990a). *Professional development guidelines for elementary school counselors: A self audit.* Alexandria, VA: Author

American School Counselor Association (1990b). *Professional development guidelines for middle/junior high school counselors: A self audit.* Alexandria, VA: Author.

Gysbers, N.C., & Henderson, P. (1994). *Developing and managing your school guidance program* (2nd ed.). Alexandria, VA: American Counseling Association

Gysbers, N.C., Lapan, R.T., Multon, K.D., & Lukin, L. (1992). *Missouri guidance competency evaluation surveys, grades 4-6, 6-9, and 9-12.* Columbia, MO: Center for Educational Assessment

Johnson, C.D. (1991). Assessing results. In S.K. Johnson & E.A. Whitfield (Eds.), *Evaluating guidance programs.* Iowa City, IA: American College Testing Program

Lapan, R.T., Gysbers, N., Hughey, K., & Arni, T.J. (1993). Evaluating a guidance and language arts unit for high school juniors. *Journal of Counseling and Development, 71*, 444-451

Starr, M., & Gysbers, N.C. (1993). *Missouri comprehensive guidance: A model for program development, implementation, and evaluation* (Rev. ed.). Jefferson City, MO: Missouri Department of Elementary and Secondary Education

Norman C.. Gysbers Ph.D. is a Professor in the Department of Educational and Counseling Psychology, University of Missouri-Columbia, Columbia, Missouri.

ERIC Digest #EDO-CG-95-7

New Assessment Methods for School Counselors

W. James Popham

The nature of classroom assessment is changing. Teachers today are being urged to rely less on traditional tests, such as those containing multiple-choice, true-false, and essay items. Instead, teachers are being encouraged to embrace innovative measurement methods, including performance tests and portfolio assessments. School counselors, if they are proactive, can help make sure that the newer assessment approaches teachers are beginning to adopt will be used in a manner that benefits students.

Many classroom teachers have never completed a formal measurement course during either their preservice or inservice classwork (Schafer & Lissitz, 1987). Not surprisingly, therefore, many teachers test their students using the same assessment procedures that they encountered during their own student days. That assessment approach is essentially a "test 'em as I was tested" strategy. It works pretty well as long as teachers are employing fairly traditional assessment methods because most teachers have been on the receiving end of more than a few of those traditional tests. But what happens when teachers try to use assessment procedures with which they have had no experience?

That is the area in which school counselors can make a meaningful contribution to the assessment acumen of the teachers with whom they work. In this digest a strategy will be described whereby student service personnel can play a leadership role in familiarizing classroom teachers and school administrators with both the payoffs and the perils of emerging classroom assessment methods.

A Special Role for Counselors

As a rule, school counselors are far more conversant with educational measurement concepts than are classroom teachers. Counselors have usually completed courses in testing (Schafer & Lissitz, 1987), and thus are not intimidated when someone talks about a validity or reliability coefficient. And all counselors know that a standard deviation is really not some sort of routine psychosis. Their familiarity with measurement procedures places counselors in a special position of perceived competence. That is, many teachers regard school counselors as experts when it comes to measurement—and that expertise is thought to include the new forms of measurement that teachers are now being urged to use. Consequently, many classroom teachers will be turning to counselors for guidance regarding the nontraditional assessment approaches they are often being told to employ. If school counselors want to make a contribution to dealing with this assessment issue, they will need to get up to speed immediately with respect to the most common of the new assessment methods, namely, performance tests and portfolios.

There are a number of books that have recently been published dealing with the innards of performance testing and portfolio assessment (e.g., Airasian, 1994; Marzano et al., 1993; Popham, 1995; Stiggins, 1994). By consulting one or more of these texts, and by focusing on their performance tests and portfolios sections, it will be possible for most counselors to acquire sufficient understanding of those two assessment approaches rapidly so that they can provide solid support for teachers. In addition, there are several digests in this special ERIC/CASS series (by Arter, by Lester and Perry, and by Stiggins) that are specifically devoted to these newer assessment approaches. Those digests provide not only useful insights regarding those assessment methods, but also identify a series of references for further reading.

To illustrate the kinds of understandings that school counselors need to acquire if they are to help their teacher colleagues deal with recent assessment advances, let us briefly consider performance tests. A performance test typically presents a task to students that calls for a relatively complex constructed response in the form of, for instance, an oral report or, perhaps, some sort of written analysis. Students' constructed responses must then be scored so that teachers can make accurate inferences about the degree to which their students possess the knowledge and/or skills assessed by the performance test.

What counselors need to know about performance assessments is: (1) how to construct the tasks for such tests; (2) how to score students' responses to those tasks; and (3) how to judge whether the performance test is a good one, that is, whether it contributes evidence that allows a teacher to make an accurate inference about a student's abilities.

Counselors should also understand the difficulty of devising and scoring such performance tests. As anyone who has scored many students' written compositions will agree, the judgment of students' composition skills is quite difficult. And yet, because we have had more than a decade's worth of experience in scoring students' writing samples, educators have worked out some fairly serviceable scoring procedures for judging students' written compositions. However, with many of the newly devised performance tasks, the difficulty of generating consistent and accurate scoring procedures is considerable. This is because of the distinctiveness of the tasks involved and, more importantly, due to our lack of experience in appraising students' responses to such tasks. Teachers need to know about such practical obstacles—and a knowledgeable counselor can inform teachers about those problems.

Let's also consider portfolios. There's much more to portfolio assessment than merely dumping a collection of student work into a manila folder. By reading about portfolio assessment, for example, counselors will learn ways of scoring the diverse student products typically found in portfolios. Counselors will also discover that many portfolio specialists believe the most significant payoff of portfolio assessment is its contribution to the student's development of self-evaluation skills. In order to foster such self-evaluation growth, the criteria for appraising portfolio products must be crisply spelled out by teachers and provided to students well in advance of the portfolio's preparation.

Portfolio conferences between the teacher and student, or even between two students, usually play a significant role in portfolio assessment strategies. Counselors will need to learn how to help teachers plan for and carry out such portfolio conferences. Counselors, obviously, need to become knowledgeable about the chief features of portfolio assessment.

The level of sophistication that a school counselor must acquire regarding portfolios and performance tests need not be off-puttingly high. Most classroom teachers do not really care about the psychometric nuances of performance tests or portfolio assessments when such schemes are employed in a statewide accountability program (e.g., Koretz et al., 1994). What classroom teachers do need to know are the nuts and bolts of performance testing and portfolio assessment as well as the strengths and weaknesses of those new assessment methods. It really should not take counselors more than a few hours of serious reading, followed by an hour or two of semi-serious thinking, to prepare themselves so they can help classroom teachers regarding these newer assessment approaches.

Whether the counselor's assistance is rendered on an individual, ad hoc basis or in a formal workshop setting will depend on the local situation. But whether a formal or informal professional development scheme is employed, a school counselor who proactively prepares to provide guidance regarding performance and portfolio assessment will clearly be in a position to supply such assistance. Counselors who do not know about portfolios or performance tests will not be able to help.

One of the eight national goals for U.S. education authorized in 1994 by the Goals 2000: Educate America Act deals with the importance of continuing professional development for teachers. Specifically, the

Goals 2000 legislation calls for teachers to remain abreast, among other things, of "emerging forms of assessment." Counselor-supplied succor regarding recent assessment advances would be most timely.

Summary

Because educators are being urged to add performance testing and portfolio assessment to their classroom assessment repertoires, many teachers will need assistance in acquiring the ability to implement such measurement techniques. School counselors can play a key role in promoting better use of these new assessment procedures if they acquire a reasonable degree of knowledge about such measurement procedures, then dispense that knowledge to the teachers with whom they work. More knowledgeable use of new classroom assessment strategies will lead to more accurate assessment-based inferences about students and, as a consequence, more defensible instructional decisions by teachers.

References

Airasian, P.W. (1994). *Classroom assessment.* New York: McGraw-Hill.

Koretz, D., Stecher B., Klein, S., & McCaffrey, D. (1994). The Vermont portfolio assessment program: Finding and implications. *Educational Measurement: Issues and Practice, 13*(3), 5-16.

Marzano, R., Pickering, D., & McTighe, J. (1993). *Assess student outcomes: Performance assessment using the dimensions of learning model.* Alexandria, VA: Association for Supervision and Curriculum Development.

Popham, W.J. (1995). *Classroom assessment: What teachers need to know.* Needham Heights, MA: Allyn & Bacon.

Schafer, W. D. & Lissitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. *Journal of Teacher Education, 38*(3), 57-63.

Stiggins, R.J. (1994). *Student-centered classroom assessment.* New York: Macmillan.

W. James Popham is an Professor Emeritus in the UCLA Graduate School of Education and Director of IOX Assessment Associates in Los Angeles.

ERIC Digest # EDO-CG-95-8

Sound Performance Assessments in the Guidance Context

Richard J. Stiggins

Not since the development of the objective paper and pencil test early in the century has an assessment method hit the American educational scene with such force as has performance assessment methodology in the 1990s. Performance assessment relies on teacher observation and professional judgment to draw inferences about student achievement. The reasons for the intense interest in an assessment methodology can be summarized as follows:

During the 1980s important new curriculum research and development efforts at school district, state, national, and university levels began to provide new insights into the complexity of some of our most valued achievement targets. We came to understand the multidimensionality of what it means to be a proficient reader, writer, and math or science problem solver, for example. With these and other enhanced visions of the complex nature of the meaning of academic success came a sense of the insufficiency of the traditional multiple choice test. Educators began to embrace the reality that some targets, like complex reasoning, skill demonstration, and product development, *require*—don't merely permit—the use of subjective, judgmental means of assessment. One simply cannot assess the ability to write well, communicate effectively in a second language, work cooperatively on a team, and complete science laboratory work in a quality manner using the traditional selected response modes of assessment.

As a result, we have witnessed a virtual stampede of teachers, administrators and educational policy makers to embrace performance assessment. In short, educators have become as obsessed with performance assessment in the 1990s as we were with the multiple choice tests for 60 years. Warnings from the assessment community (Dunbar, Kortez, and Hoover, 1991) about the potential dangers of invalidity and unreliability of carelessly developed subjective assessments not only have often gone unheeded, but by and large they have gone unheard.

Now that we are a decade into the performance assessment movement, however, some of those quality control lessons have begun to take hold. Assessment specialists have begun to articulate in terms that practitioners can understand the rules of evidence for the development and use of high quality performance assessments (e.g., Messick, 1994). As a result, we are well into a national program of research and development that builds upon an ever clearer vision of the critical elements of sound assessments to produce ever better assessments (Wiggins, 1993).

The purpose of this digest is to provide a summary of those attributes of sound assessments and the rules of evidence for using them well. The various ways the reader might take advantage of this information also are detailed.

The Basic Methodology

The basic ingredients of a performance assessment may be described in three parts (Stiggins, 1994): (1) the specification of a performance to be evaluated, (2) the development of exercises or tasks used to elicit that performance, and (3) the design of a scoring and recording scheme for results. Each contains sub-elements within it.

For example, in defining the performance to be evaluated, assessment developers must decide where or how evidence of academic proficiency will manifest itself. Is the examinee to demonstrate the ability to reason effectively, carry out other skills proficiently, or create a tangible product? Next, the developer must analyze skills or products to identify performance criteria upon which to judge achievement. This requires the identification of the critical elements of performance that come together to make it sound or effective. In addition, performance assessors must define each criterion and articulate the range of achievement that any particular examinee's work might reflect, from outstanding to very poor performance. And finally, users can contribute immensely to student academic development by finding examples of student achievement that illustrate those different levels of proficiency.

Once performance is defined, strategies must be devised for sampling student work so skills or products can be observed and evaluated. Examinees might be presented with structured exercises to which they must respond. Or the examiner might unobtrusively or opportunistically watch performers during naturally occurring classroom work in order to derive evidence of proficiency. When structured exercises are used to elicit performance, they must spell out a clear and complete set of performance responsibilities for examinees. In addition, the examiner must include in the assessment enough exercises to sample the array of performance possibilities in a representative manner that is large enough to lead to confident generalizations about examinee proficiency.

And finally, once the desired performance is described and exercises have been devised, procedures must be spelled out for making and recording judgments. These scoring schemes, sometimes called rubrics, help the evaluator translate judgments of proficiency into ratings. The assessment developer must select the level of detail to be reflected in records, the method of recording results, and who will be the observer and rater of performance.

Sound Performance Criteria

Quellmalz (1991) offers a set of specific guidelines for the development of quality performance criteria. These reflect important aspects of skill demonstration that judges are to look for and evaluate—they represent important attributes of quality products.They are devised through a thoughtful analysis of samples of high quality performance and comparison to samples of inferior performance. Out of this comparison comes an understanding of the keys to academic success in the context for which the assessment is designed. Quellmalz advises us that criteria should: be significant, specifying important performance components; represent standards that would apply naturally to determine the quality of performance when it typically occurs; be generalizable—that is, applicable to a class or tasks—not apply to only one task appropriate continuum from low- to high-level achievement; communicate clearly to and be able to be understood by all involved in the performance assessment process, including teachers, students, parents, and community; hold the promise of communicating information about performance quality that provides a basis for the improvement of that performance (p. 320).

The attributes of quality performance that form the basis of judgment criteria should be couched in the best current thinking about the keys to academic success as defined in the professional literature of the discipline in question.

Sound Performance Exercises

Baron (1991) provides guidance in the development of sound exercises. These spell out the achievement to be demonstrated by the examinee, the conditions under which the demonstrations will take place, and the criteria that will serve as the basis for evaluation of performance. In short, they focus the examinee sharply on the task at hand. Baron advises that these questions be used to determine exercise quality: when students prepare for my assessment tasks and I structure my curriculum and pedagogy to enable them to be successful on these tasks, do I feel assured that they will be making progress toward becoming genuine or authentic readers, mathematicians, writers, historians, problem solvers, etc.; do my tasks clearly communicate my standards and expectations to my students; are some of my tasks rich and integrative, requiring students to make connections and forge relationships among various aspects of the curriculum; do some of my tasks require that my students sustain their efforts over a period of time (perhaps even an entire term!) to succeed; do my tasks require self-assessment and reflection on the part of students; are my tasks likely to have personal meaning and value to my students; and do some of my tasks provide problems that are situated in real-world contexts and are they appropriate for the age group solving them?

Effective Scoring and Recording

The basis of the effective application of performance assessment methodology is thoroughly trained raters relying on sound performance criteria to observe and evaluate student responses to quality exercises (Stiggins, 1994). It is rarely the case that raters can automatically judge student performance merely as a matter of their prior professional development. Training—or at least a systematic verification of qualifications to rate performance—is essential in all contexts in which quality assessment results are the goal.

One test of the quality of ratings is interrater agreement. A high level of degree of agreement is indicative of objectivity of ratings. Another test of quality is consistency in a particular rater's judgments over time. Ratings should not drift but rather should remain anchored to carefully defined points on the scoring scale. A third index of performance rating quality is consistency in ratings across exercises intended to be reflective of the same performance—an index of internal consistency. When these standards are met, it becomes possible to take advantage of the immense power of this kind of assessment to muster concrete evidence of improvement in student performance over time.

There are three design decisions to be made by the performance assessment developer with respect to scoring schemes: the level of specificity of scoring, the selection of the record keeping method, and the identification of the rater. Scores can be holistic or analytical, considering criteria together as a whole or separately. The choice is a function of the assessment purpose. Purposes like diagnosing weaknesses in student performance that require a high resolution microscope require analytical scoring.

Recording system alternatives include checklists of attributes present or absent in performance, rating scales reflecting a range in performance quality, anecdotal records that describe performance, or mental record keeping. Each offers advantages and disadvantages depending on the specific assessment context.

Raters of performance can include the teacher, another expert, students as evaluators of each other's performance, or students as evaluators of their own performance. Again, the rater of choice is a function of context. However, it has become clear that performance assessment represents a powerful teaching tool when students play roles in devising criteria, learning to apply those criteria, devising exercises, and using assessment results to plan for the improvement of their own performance—all under the leadership of their teacher.

Performance Assessment in the Guidance Context

The ongoing guidance and counseling function in the school could bring student service personnel into contact with performance assessment methodology in three important ways. Very often, other education professionals regard counselors as sources of expertise in assessment and may bring requests for opinions about the value of this methodology, or they may ask for help in the design and development of performance assessments.

Or counselors might be invited to serve as raters of student performance in specific academic disciplines. If and when such opportunities arise, thorough training is essential for all who are to serve in this capacity. If the teachers issuing this invitation have developed or gleaned from their professional literature refined visions of the meaning of academic success, have transformed them into quality criteria, and provide quality training for all who are to observe and evaluate student performance, this can be a very rewarding professional experience. If these standards are not met, it is wise to urge (and perhaps help with) a redevelopment of the assessment.

The third and final contact for counselors is as an evaluator of students within the context of the guidance function, observing and judging academic or affective student characteristics. In this case, the counselor will be both the developer and user of the assessment and must know how to adhere to the above mentioned standards of assessment quality.

References

Baron, J.B. (1991). Strategies for the development of effective performance assessment exercises. *Applied Measurement in Education*, *4*(4), 305-318.

Dunbar, S.B., Kortez, D.M., & Hoover, H.D. (1991). Quality control in the development and use of performance assessments. *Applied Measurement in Education*, *4*(4), 289-304.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, *23*(2), 13-23.

Quellmalz, E.S. (1991). Developing criteria for performance assessments: The missing link. *Applied Measurement in Education, 4*(4), 319-332.

Stiggins, R.J. (1994). *Student-centered classroom assessment.* Columbus, OH: Macmillan.

Wiggins, G.P. (1993). *Assessing student performance.* San Francisco, CA: Jossey-Bass.

Rick Stiggins is Director of the Assessment Training Institute in Portland, Oregon.

ERIC Diget # EDO-CG-95-9

Portfolios For Assessment And Instruction

Judith A. Arter, Vickie Spandel & Ruth Culham

Portfolios are scarcely a new concept, but renewed interest, fueled by the portfolio's perceived promise for both improving assessment and motivating and involving students in their own learning, has recently increased their visibility and use. The definition of a portfolio varies some, but there seems to be a general consensus that a portfolio is a purposeful collection of student work that tells the story of student achievement or growth. (Portfolios are not folders of all the work a student does.) Within this limited definition there are portfolio systems that promote student self-assessment and control of learning; support student-led parent conferences; select students into special programs; certify student competence; grant alternative credit; demonstrate to employers certain skills and abilities; build student self-confidence; and evaluate curriculum and instruction.

Because there is no single correct way to "do" portfolios, and because they appear to be used for so many things, developing a portfolio system can spell confusion and stress, much coming from not realizing that portfolios are a means to an end and not an end in themselves. More specifically, confusion occurs to the extent there is lack of clarity on: (a) the purpose to be served by the portfolio, and (b) the specific skills to be developed or assessed by the portfolio.

It is important to keep in mind that there are really only two basic reasons for doing portfolios—assessment or instruction. Assessment uses relate to keeping track of what students know and can do. Instructional uses relate to promoting learning—students learn something from assembling the portfolio.

Instructional Uses

The perceived benefit for instruction is that the process of assembling a portfolio can help develop student self-reflection, critical thinking, responsibility for learning, and content area skills and knowledge. (It is important to point out that most of the evidence to support these claims comes from logical argument and anecdotes. There exists very little "hard" evidence that demonstrates the impact of portfolios on students.)

These benefits aren't automatic; they have to be built into the portfolio system. Suppose you are a teacher of writing. You want students to improve their ability to write and become skilled self-assessors to improve their writing. Using portfolios, what things would need to be in place? First, students need time and instruction in writing. But in addition, you and they need a clear and explicit vision of what it means to write well. How can students become skilled self-assessors if they don't know the target at which they are aiming?

This vision is often expressed using criteria that define writing performance across a range of proficiency levels. Clear criteria might specify, for instance, that a strong piece of writing would have elaborated ideas, be rich with vivid details; or have an introduction that draws the reader in while setting up what is to follow; or use an engaging, expressive voice. These criteria, which describe what it means to write well, not only serve as a guide to revision, but they provide students with a vocabulary for thinking, talking, and writing about writing. Students who internalized these criteria could use them to revise their work, reflect on it, and set goals. The students could then use a portfolio to create a collection of best writing, or diverse writing (poetry, exposition, persuasive essays, journalism, stories), or a process portfolio showing how one piece evolved from brainstorming through publication, or a growth portfolio showing how their revision skills had improved.

Ironically, the instructional benefits of portfolios are not dependent on the portfolios. Close examination of work, comparison over time, identification of strengths and weaknesses through good criteria that define quality, goal setting, connecting personal best or favorite work with who students are becoming as learners: all can occur when the vision for success is clearly defined. What is really important is not the portfolio itself so much as what students learn by creating it. Students can review and reflect on their work regularly whether or not they make a portfolio. The portfolio is a means to the end, not the end itself.

A classic example of an instructional portfolio system is the Arts PROPEL secondary creative writing, visual arts, and music portfolios in Pittsburgh Public Schools. The goals are to increase achievement levels and have students take control of their own learning through systematic reflection on work and goal setting. (See Yancey, 1992; Camp, 1992; and ASCD, 1992, for additional discussion of instructional uses.)

Assessment Uses

The perceived benefits for assessment are that the collection of multiple samples of student work over time enables us to (a) get a broader, more in-depth look at what students know and can do; (b) base assessment on more "authentic" work; (c) have a supplement or alternative to report cards and standardized tests; and (d) have a better way to communicate student progress to parents. Large-scale assessment (assessment outside of and across classrooms) tends to focus on reasons (a) and (b). Teachers tend to like portfolios for reasons (c) and (d). We will look at three common assessment uses of portfolios and then discuss some assessment issues.

Certification of Competence. A "passportfolio" shows readiness to move on to a new level of work or employment. For example, the Science Portfolio is an optional part of the Golden State Examination (California Department of Education, 1994), a large-scale assessment for high school students. It is produced during a year of science and contains a "problem solving investigation," a "creative expression" (presenting a scientific idea in a unique and original manner), a "growth through writing" that demonstrates progress in understanding a scientific concept over time, and self-reflection that enlarges on the entries. Performance criteria have been developed to judge each type of entry.

A higher stakes large-scale example is associated with "Certificate of Mastery" efforts in several states. Plans in Oregon call for portfolios to illustrate student progress toward (in the lower grades) or mastery of (by about grade 10) the state's eleven major goals for students.

Tracking Growth Over Time. A growth portfolio is a chronological collection that shows how skills, attitudes, etc., have changed over time. Early works are contrasted with later pieces. A large-scale example comes from Juneau, Alaska—The Integrated Language Arts Portfolio used in the primary grades. The portfolio is designed to replace report cards and standardized tests as ways to demonstrate growth and achievement. Growth is tracked using "developmental continuums," which describe stages of development for reading, writing, speaking, and listening. Student status on the continuum is marked at several designated times during the school year. Teacher judgments of developmental stage are backed up with samples of student work.

Accountability. Accountability uses relate to demonstrating to the community the impact of the education. A large-scale example is Vermont's grade 4 and 8 math portfolios. Students place 5 to 7 items in their portfolio to demonstrate their competence as problem solvers. The work is assessed using performance criteria for problem solving and math communication. An example at the classroom level is student-led parent conferences in which students prepare portfolios in order to demonstrate to parents what they have learned. (See Little & Allen, 1988, for an example.)

Assessment Issues. Assessment uses of portfolios, especially large-scale, high-stakes uses (for example, high school graduation), are not without controversy. Some of these issues are: 1.What is the extent to which we need to "standardize" the portfolio process, content, and performance criteria so that results are comparable? 2. Is it feasible to accurately and consistently assess student skills through portfolios? Won't this be costly? (Rand Corporation's 1992 study of the Vermont portfolio system provides an intriguing analysis of this issue.) 3. How do we get teacher buy-in? After all, teachers will be responsible for making sure that portfolios get assembled properly. 4. Will the conclusions we draw about students from their portfolios be valid? The work may not really be the students' best, or may be someone else's entirely. There are, as yet, no definitive answers to these questions, although many fear that high-stakes uses of portfolios will destroy their instructional usefulness.

Consensual Points of View

There appear to be several points on which most people agree:

Portfolios are a means to an end, not an end in themselves. The user must have a clear vision of what the "end" is.

Purpose will influence all other design and use decisions. Consider the two major purposes examined above. Portfolio systems that have assessment as the primary purpose tend to be more structured (there is more uniformity as to the items that are placed in the portfolio and the times at which they are entered); develop performance criteria primarily to allow "raters" to judge student status and monitor student growth; result in portfolios that belong to the institution; use self-reflection to gain insight about student achievement and progress; and require more time and skills for teachers to manage. Portfolios that are used for instruction tend to belong more to the student; be less structured; develop performance criteria for use by students for self-reflection; treat student self-reflection as essential for learning; and require more time and skills for students to manage. Once the purpose is clear, questions about what goes in, who decides, use of criteria, and how self-reflection is used are much easier and more logical.

There must be a clear vision of achievement targets for students. Ask this important question: What is my vision of success for my students? If you can answer this question *very clearly* you will find the process of creating portfolios much easier.

There must be student involvement in the portfolio process. Student involvement includes selecting portfolio content, developing criteria for success, and self-reflection. Even those portfolios closest to the "assessment" end of the continuum recognize the benefit from involving students in the process. If teachers put portfolios together for students, not only is this a tremendous burden for them, students learn nothing from the process. Some authors even take the position that if any other use takes precedence over instruction, portfolios will fall victim tothe same issues as past large-scale assessment attempts.

Clear and complete performance criteria are essential. For assessment purposes, we use criteria to generate scores or grades for students. However, the major value of criteria is that they assist us to articulate a clear vision of our goals for students and a vocabulary for communicating with students about these targets. Students could be partners in their development.

Conclusion

Strong portfolio systems are characterized by a clear vision of the student skills to be addressed, student involvement in selecting what goes into the portfolio, and use of criteria to define quality performance. They provide a basis for communication and self-reflection through which students share what they think and feel about their work, their learning environment, and themselves.

References

Arter, J.A. (1994). *Performance criteria, the heart of the matter.* (Available from: Northwest Regional Educational Laboratory, 101 S.W. Main, Suite 500, Portland, Oregon 97204.)

Association for Supervision and Curriculum Development (1992). *Redesigning assessment: Portfolios.* Alexandria, Virginia: ASCD.

California Department of Education (1994). *Golden state examination science portfolio.* Sacramento, California: California Department of Education.

Calkins, A. (1993). *Juneau integrated language arts portfolio for grade 1.* (Available from Juneau Borough Schools, 10014 Crazy Horse Dr., Juneau, AK 99801).

Camp, R. (1992). Portfolio reflections in middle and secondary school classrooms. In. K.B. Yancey (Ed.), *Portfolios in the writing classroom* (pp. 61-79). Urbana, IL: National Council of Teachers of English

Koretz, D., Stecher, B., & Deibert, E. (1992). *The reliability of scores from the 1992 Vermont portfolio assessment program.* (Available from: RAND Institute on Education and Training, CRESST, UCLA Graduate School of Education, 10880 Wilshire Blvd., Los Angeles, California 90024.)

Little, N. & Allan, J. (1988). *Student-led parent conferences.* (Available from: Lugus Productions Limited, 48 Falcon Street, Toronto, Ontario, Canada M4S 2P5.)

Vermont Mathematics Portfolio Project (1991). *Teacher's guide.* (Available from: Vermont Department of Education, 120 State Street, Montpelier, Vermont 05602.)

Yancey, K.B. (1992). *Portfolios in the writing classroom.* Urbana, IL: National Council of Teachers of English.

Authors

Judith A. Arter, Ph.D., directs the assessment unit at the Northwest Regional Educational Laboratory in Portland, Oregon.

Vicki, Spandel, M.A., is a senior research associate in classroom assessment at the Northwest Regional Educational Laboratory in Portland, Oregon and adjunct faculty member at Seattle Pacific University in Seattle, Washington.

Ruth Culham is a research associate in classroom assessment at the Northwest Regional Educational Laboratory in Portland, Oregon, and a former high school English teacher.

ERIC Digest # EDO-CG-95-10

Emerging Student Assessment Systems for School Reform

Edward Roeber

Currently, much discussion is taking place about the quality of American schools, the skills needed by students, and the ways we should be assessing these achievements. Student assessment is viewed nationally as the pivotal piece around which school reform and improvement in the nation's schools turn. For example, student assessment is the key piece of Goals 2000, as well as other federal legislation such as the Elementary and Secondary Education Act (ESEA).

The result is that substantially more assessment is likely to occur in our nation's schools, and to take place in areas traditionally not assessed (such as the arts), using assessment strategies (such as performance assessments and portfolios) not typically used. States and local districts are reconsidering the models for systems of assessment and how assessment at the state and local levels can be coordinated to achieve the reforms desired in education.

Why Is School Reform Occurring?

Widespread belief that schools are not helping all students achieve at the levels that are needed has spurred efforts to reform our schools. Concerns have been raised that the ways we teach students, as well as assess them, do not lead students to acquire needed knowledge or skills, nor help them apply and use their knowledge and skills appropriately. At the national and state levels, content standards containing the types of knowledge, skills, and behaviors now believed needed for all students to achieve at high levels are being developed. Starting with such efforts as the National Council of Teachers of Mathematics' Curriculum and Evaluation Standards for School Mathematics (NCTM, 1989), content standards are being developed in the arts, civics, economics, English, foreign languages, geography, health education, history, physical education, science, and social studies.

School reform is also motivated by the belief that there are competencies needed for graduates to enter the workforce successfully. The Secretary's Commission on Achieving Necessary Skills developed generic competencies and foundation skills that all workers will need in the future (U.S. Department of Labor, 1991). They include flexible problem solving, respecting the desires of the customer, working well on teams, taking responsibility for one's own performance, and continuous learning and have been developed to guide the efforts of educational reform in the direction helping more students to make the transition to work successfully.

Collectively, these standards represent substantial challenges for the American schools. They imply that all students will need to achieve at much higher levels. New strategies for assessment are also implied by these content standards.

How Does Reform of Assessment Fit School Reform?

Student assessment is at the top of the list of things to tinker with by policy makers at the national and state levels, since it is viewed as a means to set more appropriate targets for students, focus staff development efforts for the nation's teachers, encourage curriculum reform, and improve instruction and instructional materials in a variety of subject matters and disciplines (Darling-Hammond & Wise, 1985). Assessment is important because it is widely believed that what gets assessed is what gets taught, and that the format of assessment influences the format of learning and teaching (O'Day & Smith, 1993). The hope of policy makers is that changes in assessment will bring about not only the needed changes in students, but also change in the ways schools are organized (Linn, 1987; Madaus, 1985). Interest in performance assessment has also been justified on the basis that using such measures will promote educational equity (National Center on Education and the Economy, 1989). Student assessment carries a heavy load these days!

Of course, outside pressure on testing programs can be ignored or resisted by local educators (Smith and Cohen, 1991). There is also ample evidence of the distortions in teaching that external testing programs can create (Shepard & Smith, 1988). Rather than encourage reform of teaching, inappropriate teaching to the test may occur (as opposed to teaching to the domain covered by the test). Rather than creating opportunities for all students to learn to high levels, even new forms of assessment may lead to tracking and limiting opportunities for some students (Darling-Hammond, 1994; Oakes, 1985.)

Assessment reform should occur along with professional development, instructional development, and other strategies designed to assure that all of the changes are mutually supported. Coordination of assessment reform at the national and state levels with assessments at the local level is also important, so that each will present a coherent view of student performance, not simply be "stuck" together.

Types of Assessments

New content standards may require different assessment methods. Among the assessment techniques now being considered are short-answer, open-ended; extended-response, open-ended; individual interviews; performance events; performance tasks in which students have extended time; projects; portfolios; observations; and anecdotal records, in addition to multiple-choice exercises. A broader repertoire of techniques is increasingly being used.

School Improvement Strategies

The information about student achievement needed at various levels of the educational system is different. Parents have different needs than teachers, who in turn, have different needs than school principals. District administrators need broader, system-wide information, while at the state level, there is concern about equity across districts and identification of state priorities. Nationally, policy makers are concerned about differences between states and how competitive American students are with their peers in other countries.

Improving student achievement can take place at each of these levels. Teachers work with an individual student in a classroom, or revamp classroom-wide instruction based on an assessment. At the school level, educators use school information to set long- and-short-range objectives and decide how to accomplish these. At the district level, educators target particular areas of the curriculum for attention. At the state level, incentives for improving instructional programs may be most important. School reform occurs at all levels of the educational system.

Useful Assessment Designs

Typically, student achievement is measured with available student test data, often using information from district or state testing programs. Information collected less formally in classrooms is not typically included in school improvement plans, even though such information could provide valuable insights into student learning.

The nature of information needs should form the basis for an assessment design. In a top-down model, policy makers develop an assessment design that meets their needs, hoping the data may be useful by persons at lower levels. An alternative is to build the assessment system needed at the local level, aggregating the information upwards to the district, state and national levels.

Another model, based on the assumption that multiple approaches will allow different users' needs to be met, is to develop a comprehensive assessment system using different assessment formats to meet different users' needs. Various assessment strategies can be implemented together at the different levels to provide for the different information needs in a coordinated, coherent manner (Darling-Hammond, 1994).

For example, local districts can adopt a portfolio system for improving instruction, while the state carries out matrix-sampling across important standards. The information collected by the state can become part of the student's portfolio, thereby strengthening the portfolio's quality. The state could also provide opportunities for teachers to learn to score the open-ended written and performance assessments, thereby enhancing teachers' capabilities of observing and rating student performances in their classrooms.

In this case, the elements of the system at the different levels build on and support the elements at other levels. It is also anticipated that information collected at the different levels can be reported in a more understandable manner, since the same standards apply in different ways.This assessment model enhances the reforms of schools so many desire.

Summary

This is indeed a time when American schools are being challenged to provide opportunities for students to achieve at much higher levels. Assessment is viewed as one of the essential elements in assisting schools to address the standards now deemed to be important in a manner that will help all students to achieve them. The major challenge for assessment is to implement these additional assessments in a coordinated manner so that the amount of assessment is supportive of the changes needed, not overly burdensome to teachers or students. Models for coordination of assessment at the state, district and classroom levels appear most promising.

References

Darling-Hammond,L. (Spring, 1994). Performance assessment and educational equity. *Harvard Educational Review, 64* (1): 5-29.

Darling-Hammond, L. and A. Wise (1985). Beyond standardization: state standards and school improvement. *Elementary School Journal*, 85, 315-36.

Linn, R. (1987). Accountability: The comparison of educational systems and the quality of test results. *Educational Policy, 1*(2): 181-198.

Madaus, G. (1985). Public policy and the testing profession—you've never had it so good? *Educational Measurement: Issues and Practices, 4* (1): 5-11.

National Center on Education and the Economy (1989). To secure our future: The federal role in education. Rochester, NY: Author.

National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author 1989.

Oakes, J. (1985). *Keeping track: How schools structure inequality.* New Haven: Yale University Press.

O'Day, J.A. and M. Smith (1993). Systemic school reform and educational opportunity. In S. Fuheman (Ed.), *Designing coherent educational policy: Improving the system.* San Francisco: Jossey-Bass, pages 250-311.

Roeber, E. (1992). *1. Developing the comprehensive assessment system: A. top down, B. bottom up, C. both, D. none of the above.* Paper presented at the 1992 Education Commission of the States Conference on Large-Scale Assessment.

Shepard, L.A. and M.L. Smith (1988). Escalating academic demand in kindergarten: Counterproductive Policies. *Elementary School Journal, 89*, 135-145.

Smith, M. and M. Cohen (September 1991). A national curriculum in the United States? *Educational Leadership, 49* (1):74-81.

U.S. Department of Labor (1991). *Secretary's commission on achieving necessary skills.* Washington, D.C.: Government Printing Office.

Edward D. Roeber is Director of Student Assessment Programs, Council of Chief State School Officers in Washington, D.C.

ERIC Digest # EDO-CG-95-11

Assessment of Abilities

Thomas F. Harrington

This digest recommends assessing all of a person's abilities, not just some. It also discusses self-report in the context of ability assessment. The term aptitude often is used also in defining ability, and sometimes these terms are used interchangeably. Ability as used here follows Anastasi's (1988) concept of "developed abilities." Her viewpoint is that "all ability tests—whether they be designed as general intelligence tests, multiple aptitude batteries, special aptitude tests, or achievement tests—measure the level of development attained by the individual in one or more abilities" (p. 413).

What Are the Major Abilities?

In 1976, Harrington and O'Shea identified 14 abilities found in U. S. Department of Labor publications and began assessing them in a self-report format. They reviewed 113 concurrent validity studies composed of vocational/technical programs, college and university majors, and employees in different jobs, and concluded that a high degree of agreement existed between the participants' self-reports on the 14 abilities and job analysts' findings of abilities required for job performance. Later, in 1992, Harrington altered the listing by adding organization and reading ability and collapsing computational with mathematical (Harrington & O'Shea, 1993). The 15 major abilities thus identified were:

- reading
- language
- numerical/mathematical
- clerical
- technical/mechanical
- spatial
- manual
- scientific
- interpersonal
- leadership
- musical/dramatic
- organizational
- persuasive
- social
- artistic

Technical ability is a broad term that integrates many mechanical abilities. Retitling this ability acknowledges past research that shows a clear gender differentiation for mechanical ability. Schools and society should address such biases for certain abilities.

Scientific ability, a hybrid involving conceptualizing, memory, and perhaps interest in the area, requires early identification because of the hierarchical way the ability is nurtured and developed within our educational system. Developing scientific ability after little exposure is more difficult for people in their late teens and early adulthood than at a chronologically earlier age. The critical dimension is a person's exposure and identification with the unique subset of skills as being doable for him or her. Self-efficacy beliefs or feelings of adequacy, both of which can be part of the ability construct, need examination.

Readers will find the above abilities in the literature but with different names. In a summary of 25 years of research, Prediger (1992) reported the same major skills, except that he identified literary rather than the musical/dramatic ability listed above. Lowman (1991) included in his literature review 11 of the above abilities as having career relevance. He did not list four of the abilities as major abilities - scientific, reading, social, and persuasive. Instead he cited intelligence as more predictive for science occupations. He wrote, "Interpersonal skills or social intelligence appears not to be a unidimensional construct" (p. 109). He set forth a taxonomy of social demands, however, that clearly differentiates interpersonal from helping skills, which require the ability to understand the behavior and feelings of others. Lowman expressed that personal factors are most important in predicting sales performance. So science, social, and persuasive domains were recognized but were not attributed as primary abilities. Reading and language were cited among the small number of factors found in the verbal factor.

Common existing tests measure six to eight of the abilities listed in the first column above. This narrow band of abilities emerged from the multi-aptitude measures, mostly developed in the late 1940s. Job analysts, on the other hand, identify many of the aptitudes listed in the second column as necessary abilities for some jobs. Unfortunately, young people are not evaluated on these abilities and educators seldom identify them for self exploration.

It should be mentioned that knowledge of an individual's ability profile may be of moot value. Hunter (1986), after reviewing hundreds of studies, wrote "... cognitive ability predicts job performance in all jobs ... including the so-called 'manual' jobs as well as 'mental' jobs" (p. 340). He continued, "Cognitive ability predicts job performance in large part because it predicts learning and job mastery. Ability is highly correlated with job knowledge and job knowledge is highly correlated with job performance" (p. 354).

If They Are Important, Why Haven't Tests of These Abilities Been Available?

The regression model, which minimizes the number of tests used in predicting success, has dominated the field of ability measurement.

Goldman (1972), among others, pointed out that multiple aptitude batteries have limited differential predictive value and they do not offer much more than an intelligence or academic ability test. He felt that multiple aptitude tests have little to offer in counseling clients in their decision making and career planning. He wrote, "The main contribution of tests in counseling is not making predictions but facilitating the clarification of self concept" (p. 219). The National Commission on Testing and Public Policy (1990) also called for the transformation of testing in America from a gatekeeper to a facilitator. The Commission stated that testing programs must change from an overreliance on objective tests to alternative forms of assessment that help people become aware of and develop their talents. With most state plans for career development calling for students to record data about their abilities, a longer list of abilities is relevant for life planning.

What Is Self-report Methodology And How Does Its Validity Compare With The Traditional Approach?

Three different self-report assessment formats have been used. One is simply a listing of abilities with definitions and directions to indicate those areas you feel are your best or strongest. A second approach is to apply a Likert scale to a group of designated abilities. For example, in comparison to others of the same age, my art ability is excellent, above average, average, below average, or poor. Another approach is, for each ability, to provide different examples of the ability's applications on which individuals rate their performance level from high to low, and subsequently these are summed to obtain a total score. Whereas most multiple ability testing situations need several hours, the time required for the above formats ranges from 10 to 45 minutes.

Self-report assessment is cheaper and less time intrusive on a school's schedule. How do the approaches compare regarding validity? Ghiselli's (1973) summary of the average validity coefficients of different kinds of aptitude tests used to predict proficiency in the eight major occupational categories of the General Occupational Classification System shows that the coefficients are typically in the .20s and rarely go above the .30s. The average validity coefficient for prediction of proficiency on the job was .22. In a review of 55 self-evaluation of ability scales, Mabe and West (1982) reported a range of correlation coefficients from -.026 to .80, with a mean coefficient of .29 (depending on the meta-analytic method used).

More recently, Westbrook, Buck, Sanford, and Wynne (1994) demonstrated that it is possible to get acceptable reliability and validity coefficients for self-ratings which approach the size of the validity coefficients reported for objective measures of ability. Their comparative measure was the *Differential Aptitude Test* (DAT). In another study based on the common criterion of self-reported abilities of employees, Harrington and Schafer (in press) compared the abilities required for jobs from *Guide for Occupational Exploration* (GOE) job analysis data with the *General Aptitude Test Battery's* (GATB) *Occupational Aptitude Patterns* (OAP). The GOE and OAP average percentages were compared in order to evaluate which was more consistent with workers' self-expressions of their abilities. Across the 51 occupations studied, 49 of the GOE averages were larger versus one in which the OAP average was larger. It was concluded that the GOE analysis data are more congruent with worker-identified job abilities than the GATB analyses.

Conclusion

Current use of self-assessment methodology taps more ability areas than existing ability or aptitude tests cover. Alternative testing approaches have been called for which enhance self-discovery and awareness. Some recent self-report studies show at least comparable validity with more traditional approaches. Some researchers are advocating the self-assessment methodology which can substantially cut loss of instructional time and cost, evaluate hard-to-assess constructs, and deliver information most people feel is useful for self-knowledge and career planning. Philosophically, the process of self-evaluation fits the belief that individuals are in the best position to assess their own abilities since they have access to a large data base on their own successes and failures in their abilities. Most misgivings about the methodology seem to center around beliefs that individuals have a tendency to be lenient and are not objective enough in their self-analysis to provide accurate self-reports.

References

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan Publishing.

Ghiselli, E. E. (1973). The validity of aptitude tests in personnel selection. *Personnel Psychology, 26*, 461-477.

Goldman, L. (1972). Tests and counseling. The marriage that failed. *Measurement and Evaluation in Guidance, 4*, 213-220.

Harrington, T. F. & O'Shea, A. J. (1993). *The Harrington-O'Shea career decision-making system, revised manual*. Circle Pines, MN: American Guidance Service.

Harrington, T. F. & Schafer, W. D. (in press). A comparison of self-reported abilities and occupational ability patterns across occupations. *Measurement and Evaluation in Counseling and Development*.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29*, 340-362.

Lowman, R. L. (1991). *The clinical practice of career assessment: Interests, abilities and personality*. Washington, DC: American Psychological Association.

Mabe, P. A. III & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67*, 280-296.

National Commission on Testing & Public Policy (1990). *From gatekeeper to gateway*. Chestnut Hill, MA: Boston College.

Prediger, D. (1992). *Identifying and assigning abilities to ACT job clusters and Holland's occupational types*. Paper presented at the convention of the American Association for Counseling and Development, Baltimore, MD.

Westbrook, B. W., Buck, R. W., Jr., Sanford, E., & Wynne, D. C. (1994). Career maturity in adolescence: Reliability and validity of self-ratings of abilities by gender and ethnicity. *Journal of Career Assessment, 2*, 125-161.

Thomas F. Harrington is Program Director of the Counseling Psychology Program at Northeastern University, Boston, MA.

ERIC Digest # EDO-CG-95-12

Interest Assessment

Jo-Ida C. Hansen

The assessment of interests through the use of interest inventories is big business in the field of testing today. Although publishers closely guard their data on the number of inventories given, an estimate of 3,000,000 administrations per year probably is conservative. The first formal assessment of interests using a published inventory occurred in 1927 with the appearance of the *Strong Vocational Interest Blank*. Since that time, the *Strong* has survived numerous revisions and continues to be a popular and widely used interest inventory.

Interests were assessed prior to 1927 using, basically, four techniques. The earliest of these techniques was *estimation*, which simply involved asking an individual to indicate her or his feelings towards an activity. Because estimates were not always accurate, individuals often were encouraged to *try out* activities as another method for assessing their interests. Obviously, try-outs could be quite time-consuming and costly, and *rating scales* and *checklists*, precursors to interest inventories, were developed to identify interests more systematically. The interest inventories that we use today differ from early checklists and ratings in that they use statistical methods to summarize responses to pools of items representing various activities and occupations (Hansen, 1984).

Definition of Interests

The definition of interests, as used by inventory developers, researchers, and counselors, typically reflects five components that may be characterized as determinants: personality, motivation or drive, expression of self-concept or identification, heritability, and environmental influences (e.g., learning and socialization; Hansen, 1990).

One of the most popular theories for describing interests and their relationship to jobs, people, and environments is that of John Holland. Holland (1985) states that both people and environments can be divided into six vocational personality types or some combination of the six types: Realistic (outdoors, mechanical), Investigative (science, math), Artistic (art, language, music), Social (helping, teaching), Enterprising (selling, business) and Conventional (details, clerical). Holland's theory has had a tremendous impact on the fields of career counseling and interest assessment, and many interest inventories include scales that measure interests related to Holland's six types.

Purpose of Interest Assessment

Interest assessment is used in a variety of applied and research settings for several different purposes. Career exploration that leads to decisions such as choosing a major, selecting a career, or making mid-career changes, probably is the most popular and frequent use of interest assessment. Within this context, college and high school counseling services are the most typical providers of interest assessment and career counseling experiences. However, employment agencies, vocational rehabilitation services, social service agencies, corporations, consulting firms, and community agencies such as the YW or YMCA also provide career counseling opportunities that incorporate interest assessment.

Researchers use objective assessments to operationalize the construct of interests in studies that investigate variables relevant to understanding the world of work. Current trends in vocational psychology research include analyses of (a) the structure of interest; (b) the relationship of interests to other psychological variables such as personality, satisfaction, and success; and (c) the role that interests play in career development.

To a lesser extent, interests are assessed for use in selection and classification evaluations. In some instances, assessed interests, which add valuable data to career choice predictions, are used even after selection to help an employee find the right position within a particular organization (Hansen, 1994).

Current Interest Assessment Inventories

Numerous inventories designed to assess interests have been published. The available choices range from those inventories that measure a small number of relatively broad interests and are self-administered and hand-scored to those that report over 200 scores and must be scored by computer (Kapes & Mastie, 1994).

The *Self-Directed Search (SDS)* and the *Unisex Edition of the ACT Interest Inventory (UNIACT)* are based on John Holland's theory of vocational personalities and assess the six types that Holland hypothesizes. The *SDS* is self-administered, self-scored, and self-interpreted while the *UNIACT* is computer scored and uses a computer-generated narrative report to relate the scores to a World-of-Work Map.

The *Vocational Interest Inventory (VII;* 8 scales), the *Career Occupational Preference System Interest Inventory (COPS;* 14 scales), the *Ohio Vocational Interest Survey (OVIS;* 23 scales), and the *Jackson Vocational Interest Survey (JVIS;* 34 scales) feature basic interest scales that are composed of homogeneous groupings of items often identified by cluster or factor analysis. With the exception of the *COPS-R* and the *JVIS*, which can be hand- or computer-scored, all of these inventories are scored by computer. Typically these inventories measure some configuration of basic interests such as mechanical activities, athletics, nature, science, military activities, mathematics, aesthetics, social service, teaching, clerical activities, religious activities, business management, persuading, selling, health, or language.

The *Campbell Interest and Skill Survey (CISS)*, the *Kuder Occupational Interest Survey (KOIS)*, the *Career Assessment Inventory (CAI)*, and the *Strong Interest Inventory (SII)* all require computer scoring and include over 100 different measures of interests. The large number of scales allows these inventories to present profiles that include: (a) global measures of interests similar to those that represent Holland's six types; (b) basic interest scales composed of homogeneous groupings of items (e.g., scales that measure an interest in mechanical activities, medical service, or selling); and (c) scales that measure the interests of specific occupational groups such as engineers, physicians, journalists, guidance counselors, buyers, and accountants.

The choice of the appropriate inventory to use with a particular population depends on factors such as their age, the purpose of the interest assessment, the amount of time available for testing and interpretation, and the funding available to purchase materials and pay for scoring. Generally, the smaller the number of scales offered by the inventory, the less expensive the materials and scoring will be.

Computers and Interest Assessment

The option now exists to use personal computers for every phase of interest assessment, including administration of the inventory, in-house scoring of the scales, production of the profile, interpretation of the results, and integration of the assessed interests into computerized career counseling sequences (Hansen & Sackett, 1993). The most important advantage of using personal computers in interest assessment is in-house scoring that eliminates the need to mail answer sheets to a scoring service for processing, thus reducing the lag between inventory administration and interpretation of the results. A second advantage is the financial savings realized through the use of interactive computerized career guidance programs. Although these programs do not eliminate the need for counselors to work with clients, computers do provide an effective mechanism for identifying and conveying routine information and data to the client.

Summary

The assessment of interests originally developed as an outgrowth of efforts in education and in industry to supplement special and general abilities information about individuals. However, the most powerful uses of interest assessment continue to be in the context of other data, such as values, reinforcers, abilities, personality, and biographical information, that captures the life experiences of an individual. As both education and industry have discovered, the integration of a variety of information including the assessment of interests, can contribute effectively to improving individual and institutional decision-making.

References

Hansen, J.C. (1984). The measurement of vocational interests: Issues and future directions. S.D. Brown & R.L. Lent (eds.). *Handbook of counseling psychology* (pp. 99-136). New York: Wiley

Hansen, J.C. (1990). Interest inventories. In S. Goldstein & M. Hersen (eds.). *Handbook of psychological assessment* (pp. 173-194). Elmsford, NY: Pergamon Press.

Hansen, J.C. (1994). The measurement of vocational interests. In M.G. Rumsey & J.H. Harris (eds.). *Personnel selection and classification* (pp. 293-316). Hillsdale, NJ: Lawrence Erlbaum.

Hansen, J.C. & Sackett, S.A. (1993). Applications of computer technology in career interventions. B. Schlosser & K.L. Moreland (eds.). *Taming technology: Issues, strategies and resources for the mental health practitioner* (pp. 79-81). Phoenix, AZ: Division of Independent Practice of the American Psychological Association.

Holland, J.L. (1985). *Making vocational decisions* (2nd edition). Englewood Cliffs, NJ: Prentice-Hall.

Kapes, J.T., & Mastie, M.M. (1994). *A counselor's guide to career assessment instruments* (3rd edition). Alexandria, VA: American Counseling Association.

Jo-Ida C. Hansen, Ph.D., is a professor in the Department of Psychology and directs the Center for Interest Measurement Research and the Ph.D. Program in Counseling Psychology at the University of Minnesota, in Minneapolis, MN.

ERIC Digest # EDO-CG-95-13

Assessment of Self-Concept

William Strein

Self-concept is one of the most popular ideas in the psychological literature. The ERIC database includes over 6,000 entries under the "self-concept" descriptor. Unfortunately, self-concept is also an illusive and often poorly defined construct. Reviews of literature have found at least 15 different "self" terms used by various authors (Strein, 1993). Terms such as "self-concept," "self-esteem," "self-worth," "self-acceptance," and so on are often used interchangeably and inconsistently, when they may relate to different ideas about how people view themselves. Accordingly, definition is the first consideration in the assessment of self-concept. Before attempting to assess self-concept, counseling practitioners or researchers must first clarify for themselves what they mean by "self-concept" and then choose a method or instrument consistent with that definition.

Global Versus Domain-Specific Models

Perhaps the most important distinction that differentiates various conceptualizations is whether self-concept is viewed as an overarching, global characteristic of the person or as a set of self-evaluations specific to different domains of behavior. The global view, sometimes conceptualized as "self-esteem" or "general self-concept," is the older and probably the more common view among counselors and therapists (Strein, 1993). Items comprising the *Rosenberg Self-Esteem Scale* (Rosenberg, 1965) capture the essence of the global self-concept idea, and continue to be used frequently in research. The *Piers-Harris Children's Self-Concept Scale* (Piers, 1984) and the *Tennessee Self Concept Scale* (Fitts, 1991), both commonly used instruments, are also rooted in the global tradition, although each also provides domain-specific scales.

In contrast to the traditional model of global self-concept, multifaceted models stress self-evaluations of specific competencies or attributes, for example, academic self-concept, physical self-concept, and so on. Although some theoretical models are hierarchical, with global self-concept at the apex, most of these models stress the distinctiveness of various self-concept facets. Extensive empirical research in developmental and educational psychology over the past 15 years has strongly supported the multifaceted view. Consistent with research findings, most published self-concept measures now emphasize domain-specific self-concepts. The clearest example of measures based on the multifaceted view is Marsh's (1992) set of scales *(Self-Description Questionnaire I, II, or III)* covering ages seven to young adult.

Methods of Self-Concept Assessment

Self-concept is inherently phenomenological, that is, it refers to the person's own view of him- or herself. In fact, one leading scholar in the field (Wylie, 1974) has argued that comparisons to external events are not particularly relevant in the assessment of self-concept. Accordingly, self-concept is almost always assessed through self-report. Four commonly used self-report methods are described below (Burns, 1979).

Rating scales are the most frequently used type of instrument. Most of the currently published instruments are of this type. Rating scales typically are composed of a set of statements to which the respondent expresses a degree of agreement or disagreement. Five- and seven-point Likert scales are common. Typical items might be "I am good at math" or "On the whole, I am satisfied with myself." Responses are then summed to form a score for a specific scale (e.g., math self-concept) or a measure of global self-concept.

Checklists involve having respondents check all of the adjectives that they believe apply to themselves. Because the adjectives have been assigned to a category, such as "self-favorability," based on either rational or empirical criteria, the person's choices can be tabulated to form a self-concept measure. Checklists provide interesting qualitative information, but have two shortcomings. First, responses are dichotomous (yes/no); there is no way for the respondent to indicate degree of agreement. Second, the categorization of the adjectives is done by an external party, without knowing what exact meaning the adjective has for the individual.

Q-sorts have been used extensively in self-concept research but are seldom used by practicing counselors because they are time-consuming and require considerable commitment from the client. In brief, the Q-sort technique involves having the person sort cards that contain self-descriptors (e.g., "I am strong") into a pre-defined number of piles ranging from "most like me" to "least like me." Typically, 100 or more cards would be used and each pile can contain only a predetermined number of cards. Both quantitative and qualitative methods can be used to evaluate the results of the sorting task.

In *free-response* methods respondents typically complete partial statements (e.g., I feel best when...). Although some sets of these sentence-completion tasks have been published formally, complete with quantitative scoring schemes, responses more frequently are evaluated qualitatively. Free-response methods are seldom used in self-concept research but have favor with many counselors because the open-ended, qualitative nature of the task lends itself to facilitating discussion with the client. The rather low reliability of such methods, however, argues against interpreting the results as a *measure* of self-concept.

Although most of the self-concept measures compare the person's response against some set of norms, one researcher (Brahm, 1981) successfully used a *criterion-referenced approach* in which the child's self-efficacy beliefs were assessed repeatedly in reference to an external criterion of accuracy. Brahm argues that this assessment approach integrates self-concept with mastery learning more effectively than does the traditional norm-referenced self-concept scale. Although this is a promising idea, it remains undeveloped.

Considerations in the Assessment of Self-Concept

Counselors or others who wish to assess self-concept must keep several considerations in mind, including demand characteristics of self-report measures, technical adequacy of the assessment procedure, and whether the assessment is being used for research or clinical purposes. Self-report measures make several requirements of the respondent (Burns, 1979). First, the person must have a sufficient level of self-awareness. Young children may lack confidence but may not be consciously aware of their own perceptions. Second, self-report measures also require substantial verbal competence, a skill that cannot be assumed. Third, even children are aware that some responses are more socially acceptable than others. The accuracy of self-reports is often decreased by this "social desirability" response tendency.

Technical quality of self-concept instruments demands serious consideration. Reliability and validity coefficients for personality tests are frequently considerably lower than for performance measures, such as those for cognitive ability. For some of the older self-concept measures internal consistency reliabilities, especially for subscales, are only in the .70 range. Some newer instruments, however, attain internal consistency coefficients in the .90s. To help in choosing a test, prospective test users should consult technical manuals and test reviews carefully before making a final choice.

Finally, most empirically scored self-concept measures were developed more for research than for clinical use. Normative samples are seldom anywhere near as useful as for tests of achievement or ability. Information relating test scores to problem behavior is virtually absent. Counselors should use scores from self-concept measures very cautiously when working with individual clients.

References

Brahm, N. (1981). *The assessment of self concepts of educational achievement by a criterion referenced approach.* (ERIC Document Reproduction Service No. ED 235 184)

Burns, R. B. (1979). *The self-concept in theory, measurement, development and behaviour.* London: Longman.

Fitts, W. H. (1991). *Tennessee Self Concept Scale, Manual.* Los Angeles: Western Psychological Services.

Marsh, H. W. (1992). *Self-Description Questionnaire II: Manual.* Macarthur, Australia: Publication Unit, Faculty of Education, University of Western Sydney.

Piers, E. V. (1984). *Piers-Harris Children's Self-Concept Scale: Revised manual.* Los Angles, CA: Western Psychological Services.

Rosenberg, M. (1965). *Society and adolescent self-image.* Princeton, NJ: Princeton University Press.

Strein, W. (1993). Advances in research on academic self-concept: Implications for school psychology. *School Psychology Review, 22*, 273-284.

Wylie, R. C. (1974). *The self-concept, Revised edition. Volume 1.* Lincoln, NW: University of Nebraska Press.

William Strein is Associate Professor and Co-Director, School Psychology Program, University of Maryland at College Park.

ERIC Digest # EDO-CG-95-14

Assessment of Temperament

Hedwig Teglasi

Temperament refers to basic dimensions of personality that are grounded in biology and explain individual differences in the developmental process rather than universal dynamics. While these dimensions show continuity over time, they are subject to change with maturation and experience. The view of behavior as a function of the organism and of the environment is basic to psychology. Accordingly, temperament serves as a mechanism to explain how individuals contribute to their own development in a given environmental context. Harmony between persons and their surroundings is produced through bi-directional interplay between inborn, temperamental attributes and external demands, supports, and circumstances.

Temperament is generally identified with: (a) the components of personality that are biological in origin (e.g., Buss & Plomin, 1984); (b) traits that are relatively stable, cross-situationally consistent, and evident throughout the age span and diverse cultures (Rothbart & Derryberry, 1981); and (c) the style (how) rather than the content (what) or purpose (why) of behavior (Thomas & Chess, 1977). In contrast, personality serves as a central organizer of behavior that influences the expression of temperamental traits. Thus, personality determines the specific content and purpose of behavior.

Temperament is currently an active area of research with documented applicability to a variety of developmental and mental health outcomes such as conscience formation, peer interaction, behavior problems, school achievement, psychopathology, and vulnerability as well as resistance to stress. Given that temperamental extremes constitute risk factors, specific temperament dimensions can be flagged as early precursors of impaired adjustment.

Although the importance of the construct is well established, unresolved conceptual issues and problems with measurement limit the applicability of this knowledge by practitioners. The many choices of dimensions identified as separate elements, how they should be combined, and their proper measurement given these choices constitute a continuing debate. Reviews of available instruments document their problems including inconsistent stability, low interrater reliability, and questions about construct validity (Slabach, et al., 1991). Nevertheless, increasing use of temperament scales calls for research to elaborate and refine conceptualizations, to develop improved measures, and to incorporate temperament constructs in theories of personality as well as in the design of prevention and intervention strategies.

What Is the Structure of Temperament?

The nine-dimensional model of Thomas & Chess (1977) has been the basis for the development of the most popular measures of temperament in the United States. The nine dimensions are: mood, approach-withdrawal, intensity, threshold, rhythmicity, distractibility, attention span, persistence, and adaptability. However, substantial overlap found among some of these dimensions has led to questions about their validity as separate constructs. Factor analyses suggest (see review by Martin, et al., in press) that these nine dimensions separate into five robust factors and two factors that are less consistent across measures and ages.The five robust factors are: inhibition (approach-avoidance), negative emotionality, adaptability, activity level, and task persistence. The two less consistent factors are: threshold and biological rhythmicity. The five robust dimensions emerging from the factor analytic study of childhood temperament resemble the Big Five factors identified in the study of adult personality and suggest a temperamental underpinning to personality.

Buss and Plomin (1984) emphasized the two criteria of early appearance and heritability as defining properties of temperamental traits and developed a measure based on the following three dimensions: emotionality, activity, and sociability (EAS). Factor analysis of a selected set of items from the EAS and the nine-dimensional model (ages 1-6) suggested the following factors: emotionality, soothability, activity, attention span, and sociability (Rowe & Plomin, 1977).

Rothbart and Derryberry (1981) defined temperament as constitutionally based individual differences in reactivity and self-regulation (influenced over time by heredity, maturation, and experience). Reactivity refers to the activation of motor, affective, autonomic, and endocrine systems. Self-regulation refers to the processes that modulate reactivity such as attention, approach-withdrawal, inhibition, and self-soothing. This framework broadens the possibility of identifying temperament dimensions to include those that do not appear within the first years of life. Furthermore, this approach promotes the application of research in areas such as emotion and cognition to refine temperament dimensions. In developing a series of temperament questionnaires for various ages, Rothbart and her colleagues identified as many as 15 dimensions of temperament, some of which are refinements of those previously identified such as emotionality (see Goldsmith & Rothbart, 1991).

What Issues Remain in Assessing Temperament?

One problem in the assessment of temperament is that measures for older children have been either upward extensions of temperament constructs and scales derived from observations on infants and toddlers or biological models without regard to development. An emphasis on early appearing traits precludes the consideration of characteristics that may be genetically programmed to emerge later in time, and the disregard of developmental processes excludes from consideration age-related variation in the expression of temperament. Developmental changes in the elicitors of temperamental responses such as fear or pleasure have been studied in the early years through contrived laboratory situations, but such prototypical situations at later ages remain to be determined.

Response parameters need to reflect the greater complexity and differentiation of behavior with development. Commonly assessed response parameters in laboratory studies with young children have been duration, latency, and intensity. However, other parameters that tap the greater organization of behavior with development might entail modulation, self-regulation, or attunement to context. Furthermore, age and rater differences in the meaning of specific items on scales have not been investigated.

How Are Temperament and Personality Related?

Despite efforts to distinguish between temperament and the more general concept of personality, the contrast between them is obscured by the following (see Prior, 1992): (a) a common descriptive vocabulary; (b) overlapping concepts; and (c) failure of empirical data to differentiate between temperament and personality on the basis of biological factors.

The concept of self-regulation, widely studied as a personality variable, has also been regarded as a temperamental trait. Self-regulation as a personality construct appears to be defined in general terms encompassing the manner in which an individual thinks, feels, acts, and reacts. The temperament view refers to the basic processes involved in optimizing stimulation, alertness, and affective arousal.

Needed is an explanation of how the basic response styles identified as temperamental traits express themselves in larger units of functioning such as self-regulation in the broader sense. Temperament contributes to the coherence of the individual's current functioning and to both continuity and lawful changes in the developmental process. The individual's current state (personality) can be framed in terms of unfolding processes (continuous interaction between person and environment) that led to its development.

How Do Temperament Dimensions Exert Their Influence?

The mechanisms by which temperament dimensions exert their influence on broader areas of functioning are less well understood than the traits themselves. Martin (1994) reviewed two possible causal linkages between temperamental dispositions and children's common problems in educational settings that focus on the interplay of temperament with the environment:

1. Some components of the environment strengthen temperamental dispositions because the environment that is actually experienced is linked with those predispositions in three ways: (a) on average, children share 50% of their own genetic makeup with each of their parents who then provide environments that are influenced by their own genetic backgrounds; (b) children's behavioral styles (i.e., temperaments) elicit responses from others in the environment in ways that strengthen their disposition; and (c) children actively seek environments that are in harmony with their predispositions.

2. Temperament acts as a predisposition to (or buffer against) risk in the context of stressful conditions. According to this model, the role of the environment varies with the degree of predispositional risk.

A third possibility, that temperament influences the perception and synthesis of life experiences, is suggested by research on the impact of emotion on information processing and memory. Similarly, attentional processes, considered by many as temperamental, would be expected to have a very basic impact on the interpretation of information. Over time, the cumulative influence of temperament on the understanding of experiences (social and task) shapes the individual's inner world including views of relationships and expectations about events. These inner structures corroborate and amplify the original predispositions. Strategies to intervene must be aimed at altering the processes set into motion by the individual's temperamental dispositions.

Conclusions

Temperament is a compelling framework within which to study the contribution of individual differences to the developmental process. The documented association of temperament traits with diverse outcomes linked with normal development and psychopathology have left no doubt about the value of this construct. Future refinements in definitions and measurement as well as a better understanding of how temperament exerts its influence will promote greater application of these concepts to designing programs for prevention and intervention in mental health and educational settings.

References

Buss, A., & Plomin, R. (1984). *Temperament: Early personality traits.* Hillsdale, N.J.: Erlbaum.

Goldsmith, H. H., & Rothbart, M. K. (1991). Contemporary instruments for assessing early temperament by questionnaire and in the laboratory. In A. Angleitner & J. Strelau (Eds.), *Explorations in temperament: International perspectives on theory and measurement.* New York: Plenum.

Martin, R. P. (1994). Child temperament and common problems in schooling: Hypotheses about causal connections. *Journal of School Psychology, 32*, 119-134.

Martin, R. P., Wisenbaker, J., & Huttunen, M. (In press). Review of factor analytic studies of temperament measures based on the Thomas-Chess Structural Model: Implications for the Big Five. In C. Halverson, Jr., G. Kohnstamm, & R. P. Martin Eds.), *The developing structure of temperament and personality from infancy to adulthood.* Hillsdale, NJ: Erlbaum.

McCrae, R. R. (Ed.) (1992). The five-factor model: Issues and applications (Special issue). *Journal of Personality, 60.*

Prior, M. (1992). Childhood temperament. *Journal of Child Psychology and Psychiatry, 33*, 249-279.

Rothbart, M. K., & Derryberry, D. (1981). Development of individual differences in temperament. In M. E. Lamb & A. L. Brown (Eds), *Advances in developmental psychology* (Vol. I, pp. 37-86). Hillsdale, N. J.: Erlbaum.

Rowe, D.C., & Plomin, R. (1977). Temperament in early childhood. *Journal of Personality Assessment, 41*, 150-156.

Thomas, A., & Chess, S. (1977). *Temperament and development.* New York: Brunner/Mazel.

Hedwig Teglasi, Ph.D., is an Associate Professor and Co-Director of the School Psychology Program at the University of Maryland, College Park, MD.

ERIC Digest # EDO-CG-95-15

Assessment of Preschool Children

Nicholas A. Vacc & Sandra H. Ritter

Assessment of Preschool Children

With the enactment of the *Education for All Handicapped Children Act* (PL 94-142) of 1975 and its amendments (PL-99-457 of 1986 and PL 101-476 of 1990), all children are entitled to appropriate free education and related services regardless of disabilities. As a result, major strides have been made toward providing services for developmentally delayed children. These services include transportation, case management, family training and counseling, home visits for counseling, health services, medical services for diagnostic purposes, nursing services, nutrition services, occupational therapy, physical therapy, psychological services, social-work services, special classroom instruction, adapted physical education, audiology, and speech-language pathology. To gain access to these services, children who are suspected of having developmental or physical disabilities have to be referred to trained and qualified individuals or multidisciplinary teams for assessment in cognitive, physical, language and speech, psychosocial, and self-help areas.

Young children, however, are difficult subjects to assess accurately because of their activity level and distractibility, shorter attention span, wariness of strangers, and inconsistent performance in unfamiliar environments. Other factors that may affect a child's performance include cultural differences and language barriers, parents not having books to read to their child and a child's lack of interaction with other children. Consequently, assessment of infants, toddlers, and young children requires sensitivity to the child's background, and knowledge of testing limitations and procedures with young children.

Current Trends

Assessment, differentiated from test administration and interpretation, is usually a comprehensive process of gathering information about a child across developmental areas. Benner (1992) reported several continua along which assessments fall: (a) norm-referenced to criterion-referenced, product oriented to process oriented assessment; (b) formal to informal assessment, direct to indirect assessment; (c) standardized tests to handicap-accommodating tests; and (d) single-discipline approach to team approach. The present trend in preschool assessment is toward the latter perspective of each continuum with strengths being emphasized rather than deficits.

Thus, current trends in preschool assessment include a move away from a "single assessor" model to an environmental model which is designed for the individual child. Through a team approach, children are evaluated with family members present, and factors of the home and social environment are taken into consideration. Because of the increased situation-specificity of developmental tests, which can be administered by professionals other than practicing psychologists, their use is increasing (Niemeyer, J. A., personal communication, August 19, 1994).

It has been recommended that norm-referenced tests, such as intelligence tests which historically have been used as a measure of ability and as an entrance criterion for programs such as Head Start, be replaced with assessments based on multiple theoretical perspectives (Niemeyer, J. A. personal communication, August 19, 1994). A more holistic evaluation of the child can be obtained by integrating tests of cognitive ability with other measures such as assessment of social and motor skills development.

Characteristics of Preschool Assessment

In identifying appropriate interventions at the preschool level, there is less focus on testing and more on evaluating the individual child. Some of the more important characteristics are as follows:

Criterion referenced and process oriented

Criterion-referenced tests allow each child to be assessed as an individual. Comparing the child with developmental milestones and selecting areas to reinforce allows interventions to be specifically tailored to a child. Attention is given to the *process* of the interactions (i.e., whether the assessment is being conducted in a way that optimizes the child's demonstration of abilities).

Informal, indirect, and naturalistic evaluations

Informal, relaxed settings where the child can be as much at ease as possible are recommended when doing assessment. Assessing a child within the context of his or her community and the interacting social systems, and taking into account the family's needs, resources, and concerns affect both the evaluation and possible interventions. One of the most important developments in this area is Transdisciplinary Playbased Assessment (Linder, 1993), during which the child engages in play with a familiar person and the interactions of the child with the adult are observed by a team. The assessment is constructed so that the team can communicate with the play facilitator concerning unobserved skills (e.g., can the child stack three blocks).

The combination of informal play-based assessment and more directed and structured activities provides greater opportunity for a high level of performance (Bagnato & Neisworth, 1994).

Handicap accommodating assessments

Standardized assessment procedures present problems when a child has a handicap that impedes test performance even though the area being examined is not related to the handicap. Attention is being directed toward developing assessment procedures that accommodate for handicaps and provide a more accurate evaluation of the child.

Multi-disciplinary/trans-disciplinary approach

Because single-discipline evaluations provide a "snapshot" from a limited perspective, assessments involving more than one discipline are recommended. Options include multi-disciplinary, inter-disciplinary, and trans-disciplinary assessments. Multi-disciplinary teams are based on the medical model where many disciplines evaluate individually and provide reports to a central figure. Inter-disciplinary team members assess the child individually and then convene to discuss findings and form joint recommendations. With a trans-disciplinary team, representation of all disciplines that are needed for a child (e.g., occupational therapy, speech therapy, medical doctor, nutritionist) are present, and the child is observed and discussed by all at the same time, thus providing an evaluation of the total child.

The Role of Mental Health Practitioners

Many current methods for preschool assessment are designed to be convenient for both the assessors and the families, and to have all individuals involved with a child participate directly in the evaluation process. Improvement is fostered when a holistic concept of the child is provided through a multi-disciplinary or trans-disciplinary assessment that is part of a larger set of conditions which promote change, such as family system interventions (AAHE, 1992). In many instances, the mental health practitioners (e.g., counselors) will not be directly involved in the test administration, but will work with the family during the process. In particular, mental health practitioners can provide information on testing, legal requirements, and the merits and limitations of preschool assessment methods. It is helpful for the parents to know that the principles of good assessment practice reflect a multi-dimensional, integrated understanding of learning, explicitly stated purposes, experiences that lead to results, and continuous intervention and re-evaluation. Mental health practitioners who are actively involved as part of the assessment team evaluating a referred child need to be familiar with the different assessment methods and their limitations, as well as current assessment trends and the reasoning behind them. This is especially important given that as few as 10% of tests administered to preschool children have been reported as appropriate to screen that population (Wortham, 1990). If mental health practitioners are not participants in the assessment process but are in the position of working with a child or the family after an assessment has been completed and a referral has been made, they need to evaluate whether the instruments employed, the assessment environment, and the way in which the evaluation was administered were appropriate for the particular child.

Summary

Major changes in the level of interest and evaluation methods employed in preschool assessment have occurred in the past decade. The current trend is toward an ecological, child-centered approach which includes trans- or multi-disciplinary evaluations. Such approaches evaluate the "total child" rather than a specific area.

References

American Association for Higher Education (1992). *Principles of good practice for assessing student learning*. Washington, DC: Author.

Bagnato, S. J. & Neisworth, J. J. (1994). A national study of the social and treatment "invalidity" of intelligence testing for early intervention. *School Psychology Quarterly, 9*(2), pp. 81-102.

Benner, S. M. (1992). *Assessing young children with special needs: An ecological perspective*. New York: Longman.

Linder, T. W. (1993). *Transdisciplinary Play-based Assessment: A functional approach to working with young children*. Baltimore, MD: Paul H. Brookes.

Wortham, S. C. (1990). *Test and measurement in early childhood education*. Columbus, OH: Merrill.

Nicholas A. Vacc, Ed.D. is Professor and Chair of the Department of Counseling and Educational Development at the University of North Carolina at Greensboro.

Sandra H. Ritter, M.Ed. is a doctoral student in the Department of Counseling and Educational Development at the University of North Carolina at Greensboro.

ERIC Digest #EDO-CG-95-16

Screening for Special Diagnoses

Susan De La Paz and Steve Graham

Congress enacted Public Law 94-142, *the Education for All Handicapped Children Act*, in November 1975. It requires that all children with disabilities receive a free and appropriate public education. Determining who has a disability and who is eligible for special services, however, is not an exact science. It is complicated by vague definitions and varying interpretations of how to identify specific handicapping conditions (Hallahan & Kauffman, 1991). Nevertheless, recent government figures indicate that 7 percent of children and youth from birth to 21 are identified as having a disability that requires special intervention (Hunt & Marshall, 1994).

While practices differ greatly both across and within states (Adelman & Taylor, 1993), screening is an important part of the assessment process mandated by Public Law 94-142. Screening for the purpose of special diagnoses begins at birth and continues throughout the school years. In the first few years of life, most forms of screening center around developmental norms for physical, cognitive, and language abilities. Many children with severe disabilities (cerebral palsy, spina bifida, Down's syndrome, autism, severe sensory impairments, or children with multiple disabilities, for example) are identified early in life by physicians and other health professionals. However, other children, such as those with learning disabilities, attention deficit disorders, behavioral problems, and so forth, are usually not identified until they start school.

School-Based Screening

Most public schools periodically "screen" large groups of students, typically between kindergarten through third grade, to identify children who may have a disability (as yet unidentified) or may be at risk for school failure. For example, a student with an extremely low test score on a standardized achievement test administered to all first graders in a school may become the focus of further inquiry to determine the validity of the screening observation and, if warranted, to determine the causes of the child's difficulties. This may lead to a recommendation to conduct a formal evaluation to decide if the child has a specific, identifiable disability. In addition to systematically "screening" students, children with a "suspected" disability may also be identified through referrals by parents, teachers, or other school personnel.

Typically, a child who is having academic or behavioral problems in the classroom may be referred for further testing to determine if a disability is present. Before testing for diagnosis begins, however, the school must obtain consent from the child's parents to do the evaluation.

While most children with a disability are identified by third grade, some are not identified until the upper elementary grades or even junior or senior high school. In some instances, a problem does not become evident until the demands of school exceed the child's skills in coping with his or her disability. In other cases, the disability may not occur until the child is older. For instance, a disability may be acquired as a result of a traumatic brain injury or as a result of other environmental factors. A disability may also not be identified until a child is older because the procedures used for screening, referral, testing, and/or identification are ineffective.

Problems and Solutions for School Screening

It is important to understand that there is no standard or uniform battery of tests, checklists, or procedures to follow for the identification of most students with disabilities. While there is a basic structure to the identification process, there is considerable variability in how students may come to be identified, including the types of tests used in screening and the processes by which they are referred.

Critics have argued that the procedures used to identify children and youth with special needs have resulted in over- as well as under-identification of students with disabilities. As several studies have shown, a referred child almost always qualifies for special education (Christenson, Ysseldyke, & Algozzine, 1982). Over-identification has been particularly problematic in the area of learning disabilities (Hunt & Marshall, 1994), as approximately half of all students receiving special education services are identified as learning disabled! In contrast, students with behavioral disorders appear to be under-identified, particularly children who are compliant and nonaggressive but suffer from problems such as depression, school phobia, or social isolation (Walker et al., 1990).

To remedy problems of over- and under-identification, educators have begun to institute several changes in the screening and referral process. One approach has involved the development of better screening procedures. For example, Walker and his colleagues (1990) devised a screening process, the Systematic Screening for Behavioral Disorders, that relies on a three-step process. Teachers (1) rank-order students along specified criteria and then (2) use checklists to quantify observations about the three highest-ranked students. Then, (3) other school personnel (for example, school psychologists or counselors) observe children whose behaviors exceed the norm for the teacher's classroom. Referrals are made for further evaluation only after the three-step process is completed.

A second common practice aimed at improving the identification process involves the use of prereferral interventions (Chalfant, 1985). These interventions have been developed to reduce the number of referrals to special education and provide additional help and advice to regular education teachers. Before initiating a referral for testing for special diagnoses, teachers first attempt to deal with a child's learning or behavioral problems by making modifications in the regular classroom. If these modifications fail to address the difficulties the child is experiencing adequately and the teacher believes that special services may be warranted, then the referral process is set into motion. Currently, 34 of 50 states require or recommend some form of prereferral intervention (Sindelar, Griffin, Smith, & Watanabe, 1992).

Two of the more common prereferral intervention approaches include Teacher Assistance Teams (TATs) and collaborative consultation. Both approaches involve professionals helping regular educators deal with students who have problems in their classroom; however, they differ in an essential way. TATs typically consist of a team of three teachers with the referring teacher as the fourth member. The TAT model provides a forum where teachers meet and brainstorm ideas for teaching or managing a student. In contrast, most collaborative consultation models employ school specialists (resource room teachers, speech-language clinicians) who work directly with the referring teacher to plan, implement, and evaluate instruction for targeted students in the regular classroom.

Summary

Screening procedures are an important part of the assessment process to identify children and youth who have disabilities. Such procedures must be used with care, however, as they provide only a preliminary sign that a child has a disability. Additional testing is required to affirm or disprove the presence of a handicapping condition. If a disability is identified during follow-up assessment, the focus shifts to providing the student with an appropriate education.

References

Adelman, H., & Taylor, L. (1993). *Learning problems and learning disabilities*. Pacific Grove, CA: Brooks.

Chalfant, N. B. (1985). Identifying learning disabled students: A summary of the National Task Force Report. *Learning Disabilities Focus, 1*, 9-20.

Christenson, S., Ysseldyke, J., & Algozzine, B. (1982). Institutional constraints and external pressures influencing referral decisions. *Psychology in the Schools, 19*, 341-345.

Hallahan, D., & Kauffman, J. (1991). *Exceptional children*. Englewood Cliffs, NJ: Prentice Hall.

Hunt, N., & Marshall, K. (1994). *Exceptional children and youth*. Boston, MA: Houghton Mifflin Company.

Mercer, C. (1991). *Students with learning disabilities*. New York, NY: Merrill.

Sindelar, P., Griffin, C., Smith, S., & Watanabe, A. (1992). Prereferral intervention: Encouraging notes on preliminary findings. *The Elementary School Journal, 92*, 245-259.

Walker, H., Severson, H., Todis, B., Block-Pedego, A., Williams, G., Haring, N., & Barckley, M. (1990). Systematic screening for behavior disorders (SSBD): Further validation, replication, and normative data. *Remedial and Special Education, 11*, 32-46.

Susan De La Paz is a Doctoral Candidate in the Department of Special Education, University of Maryland, College Park.

Steve Graham is Professor, Department of Special Education, University of Maryland, College Park.

ERIC Digest # EDO-CG-95-17

Assessment in Career Counseling

Dale J. Prediger

"In choosing an occupation one is, in effect, choosing a means of implementing a self-concept" (Super, 1957, p. 196). What might be called "Super's Dictum" has an antecedent in ancient Greek thought: "Know Thyself." It was formulated in the early days of the career development revolution that eventually swept away square-peg-square-hole thinking about assessment. Current thinking regarding the role of assessment in career development and counseling represents an extension of Super's Dictum and a revitalization of trait and factor theory.

Since the content of assessment in career counseling (e.g., interests, abilities, career certainty) is well-covered by other digests in this series (also see Kapes, Mastie, & Whitfield, 1994), this digest focuses on the process—specifically, the contribution of assessment procedures to career exploration and planning. (Super's Dictum on choosing an "occupation" encompasses the *trial occupational choices* characterizing exploration and planning.) Because these career development tasks are experienced by everyone, this digest addresses assessment for the many (e.g., via career planning courses) rather than intensive, problem-focused career counseling.

Basic Considerations

1. Trait and Factor Theory: The Foundation for Assessment

Assessment procedures used in career counseling have their roots in tests used for diagnostic screening and personnel selection (hiring). As a result, the "test 'em and tell 'em" approach to test use and the focus of scores on arbitrary decision points (e.g., helping Pat choose a career at 10:20 a.m. on Tuesday, March 17th) were major problems at one time. Trait and factor theory was and continues to be blamed for these problems. However, there is nothing inherently wrong with assessing human traits. Indeed, assessment is part of human nature; for millennia, we have "sized-up" strangers and acquaintances. Misinterpretations and misapplications of trait and factor theory are now widely recognized and there have been several recent attempts to place trait assessment into the context of career development theory (e.g., see Chartrand, 1991; Rounds & Tracey, 1990).

2. Self-Concept: The Basis for Career Choice

According to Super's Dictum, an occupation gives one the chance to be the kind of person one wants to be; hence, career choices are based on self-concepts projected into career options. It follows that a major task in career counseling is to elicit and inform self-concepts—not a simple process (Betz, 1994) unless one prioritizes components according to career relevance. Faulty self-concepts are likely to result in flawed plans and choices. Herr and Cramer (1992) said it this way: "The major concern in a career [development] model is the clarity and accuracy of the self-concept as the evaluative base by which to judge available career options" (p. 155).

3. *Assessment: A Primary Means for Self/Career Exploration*

Given today's complex array of career options, one of the most difficult developmental tasks persons face is the identification and exploration of options congruent with their characteristics. Assessment can provide focus to career exploration. In the process of assessment and career exploration, counselees will develop insights about themselves and the work world that will inform their self-concepts. In a nutshell, the major role of assessment in career counseling is *self/career exploration*—a complementary process.

4. *Transformation of Assessment Data: Requirement of Helpful Assessment*

Assessment data (standard scores, percentile ranks, etc.) must go through a series of transformations if they are to be helpful in career counseling. First, data must be transformed into counseling information—i.e., career options worthy of exploration. Next, a short list of options must be transformed into action—i.e., self-evaluated activities and experiences. Finally, self-evaluations and self-concepts must be transformed into career plans. Because of the research and technology involved (see below), counselors should require that test publishers take primary responsibility for the first transformation. Counselors and counselees share responsibility for the other two.

5. *Data-Information Transformation: Bridge to Reality*

In a 30-year-old text on test interpretation fundamentals (many of which are ignored today), Goldman (1971) described the following three models for transforming assessment data into counseling information—for "bridging the gap" between a score and its real-world implications.

Clinical interpretations: Bridge for those with time. The labor-intensive clinical interpretation model (see Goldman, 1971, for specifics) is shaky at best—unless counselors are very well trained and have a light load. It is often supported by little more than a list of scores; a vague understanding of measurement error, "validity coefficients," and "hit rates"; specific knowledge about a few occupations and a mystical reliance on counselor/counselee intuition. While intuition can contribute to assessment for career counseling, counselors should expect publishers of assessment instruments to help them "bridge the gap" between scores and their implications.

Success predictions: Bridge to nowhere. Presumably, the prediction model can forecast levels of occupational success. Presumably, a counselor can say (for example): "Pat, based on your test scores, chances are about 59 out of 100 that you will be moderately 'successful' as a counseling psychologist and 27 out of 100 that you will be highly 'successful.' Now, as for flight attendant and pediatrician, . . ." Unfortunately, research indicates that so-called "actuarial methods" can never provide predictions of occupational success for enough occupations and with enough precision to be of use in career exploration (e.g., see Goldman, 1994; Prediger, 1974). Nevertheless, the latest claim is that success predictions based on general mental ability (formerly called IQ) can be provided *and compared* across nearly all occupations. This is despite the facts that: (a) "success" is defined differently from occupation to occupation (b) defensible measures of level of success are often unavailable (e.g., for counseling psychologist, pediatrician); (c) predictor-success correlations are available for relatively few occupations; and (d) when available, prediction errors are large.

Attempts to predict occupational choice are also unwarranted. Besides, what counselor would want to say (for example): "Pat, chances are 73 out of 100 you will become a nurse, [etc.]"? According to Zytowski (1994), the prediction model "is the failed relationship" (p. 222) between tests and career counseling.

Similarity estimates: Bridge to the work world. The similarity model ("you look like a person who") can be used to survey the work world in order to identify occupational options warranting exploration. (For over 60 years, interpretations of the Strong Interest Inventory Occupational Scales have been based on this model.) The goal of the similarity model is *not* to predict level of success or to find the "ideal career." Rather, the goal is to say (for example): "Pat, here are some occupations that attract people who are similar to you in several important ways. You may want to check them out." Research shows that observed differences among career groups are of sufficient magnitude to provide focus to career exploration (e.g., see Prediger, 1974; Rounds & Tracey, 1990; Zytowski, 1994). Counselors should expect publishers of assessment instruments to provide them with an interpretive bridge based on similarity model research. Improvised, armchair "structured searches" should be questioned.

6. Informed Self-Estimates: Key to Ability Assessment

Unfortunately, test scores are seldom available for many work-relevant abilities—e.g., sales, leadership/management, organization, creative/artistic, social interaction. Too often, work-relevant abilities that can't be assessed by paper-and-pencil tests are ignored. But career exploration based only on abilities for which there are tests not only misses important abilities, it does not take account of the powerful role of self-concepts in occupational choice (recall Super's Dictum). Ability self-estimates bring work-relevant self-concepts to the attention of the counselee and the counselor. Elsewhere, I have discussed how *informed* self-estimates of abilities can be used to facilitate self/career exploration (Prediger, 1994). To be accurate, self-estimates must be informed by experience—including the ability *estimates* provided by test scores, if they are available.

7. Comprehensive, Articulated Assessment: A Goal

Career development theory makes it hard to defend career exploration based only on interests, only on abilities, or only on job values (e.g., see Lowman, 1993). Nevertheless, some counselors still take a piecemeal approach to career assessment—e.g., interests in Grade 9; abilities 3 years later. Counselors may also face the problem of interpreting interest, ability, etc., assessments based on different norms, profile formats, and work world structures. Some publishers are responding to these problems with comprehensive, articulated assessment programs. Counselors should expect nothing less.

8. Development of Possibilities into Realities: A Requirement

One of the career counselor's primary functions is to help counselees develop career possibilities into realities—that is, to facilitate personal growth (e.g., building the abilities needed for a preferred career path). In conjunction with other information about the counselee, assessment information can suggest where growth would be helpful and how it can be effected.

Summary

Trait and factor theory,(now "person-environment fit theory") has been revitalized by career development theory. Recognition of the importance of the self-concept in career exploration provides the basis for a closer relationship between assessment and counseling.

References

Betz, N. E. (1994). Self-concept theory in career development and counseling. *Career Development Quarterly, 43*, 32-42.

Chartrand, J. M. (1991). The evolution of trait-and-factor career counseling: A person x environment fit approach. *Journal of Counseling & Development, 69*, 518-524.

Goldman, L. (1971). *Using tests in counseling* (2nd ed.). New York: Appleton-Century-Crofts.

Goldman, L. (1994). The marriage is over. . . for most of us. *Measurement and Evaluation in Counseling and Development, 26*, 217-218.

Herr, E. L., & Cramer, S. H. (1992). *Career guidance and counseling through the life span* (4th ed.). New York: HarperCollins.

Kapes, J. T., Mastie, M.M., & Whitfield, E.A. (In press). A counselor's guide to career assessment instruments. Alexandria, VA: National Career Development Association.

Kapes, J.T., Mastie, M. M., & Whitfield, E.A. (Eds. 1994). *A counselor's guide to career assessment instruments* (3rd ed.). Alexandria, VA: National Career Development Association.

Lowman, R. L. (1993). The inter-domain model of career assessment and counseling. *Journal of Counseling and Development, 71*, 549-554.

Prediger, D. J. (1974). The role of assessment in career guidance. In E. L. Herr (Ed.), *Vocational guidance and human development* (pp. 325-349). Boston, MA: Houghton Mifflin.

Prediger, D. J. (1994). Tests and counseling: The marriage that prevailed. *Measurement and Evaluation in Counseling and Development, 26*, 227-234.

Rounds, J. B., & Tracey, T. J. (1990). From trait-and-factor to person-environment fit counseling: Theory and process. In W. B. Walsh & S. J. Osipow (Eds.), *Career counseling: Contemporary topics in vocational psychology* (pp. 1-44). Hillsdale, NJ: Erlbaum.

Super, D. E. (1957). *The psychology of careers*. New York: Harper & Brothers.

Zytowski, D. G. (1994). Tests and counseling: We are still married, and living in discriminant analysis. *Measurement and Evaluation in Counseling and Development, 26,* 219-223.

Dale J. Prediger, Ed.D., NCCC, is a Vocational Research Psychologist with American College Testing, Iowa City, IA.

ERIC Digest # EDO-CG-95-18

Assessing Career Certainty and Choice Status

Paul J. Hartung

Career certainty refers to the degree to which individuals feel confident, or decided, about their occupational plans. The construct proves elusive to explain clearly unless considered in terms of the larger domain of career decision making and, specifically, career indecision. Research has yielded a variety of instruments useful for assessing career indecision. These instruments typically include a measure of career certainty by using one or two items that in part comprise a larger inventory that surveys career choice status. These measures give counselors important practical tools for appraising clients' career choice status as a step in assisting clients to alleviate their career indecision. Measures of career certainty and indecision also provide researchers with a means of determining the efficacy of career counseling interventions.

Parsons (1909) pioneered the study and assessment of career certainty and career indecision. In his work, he classified people into career-decided (i.e., certain) and career-undecided (i.e., uncertain) groups. Some years later, Williamson (1937) discounted empirically the then widely held belief that certainty of vocational choice predicts scholastic achievement. As part of his research, Williamson asked students reporting definite vocational choices to rate themselves as very certain, certain, or uncertain about their choices. Research such as Williamson's that used Parsons' dichotomous model to study career-decided and career-undecided groups produced mixed and inconsistent results. Some studies found that decided and undecided people showed significant personality or performance differences, whereas other studies found no differences between these two groups (see Slaney, 1988a for a review). As one way of resolving these inconsistent findings, researchers reconceptualized undecided people as comprising different sub-types and turned to developing psychometric instruments that would assess degree of and reasons for career uncertainty. Work by Savickas (1992) suggests that these measures now constitute two generations of instrument development.

First-Generation Measures

First-generation measures of career choice status yield total indecision scores. Such instruments, although not multidimensional by design, have generated considerable research on identifying multiple subtypes of undecided people and developing differential interventions for each type.

Initially called the Types Questionnaire, the Career Decision Scale (CDS; Osipow, Carney, Winer, Yanico, & Koschier, 1976) ranks as the prototypical first-generation measure. The original title of the CDS reflected the purpose of the instrument to scale various decisional problem types and to measure antecedents of career indecision. Although predated by other measures, such as the Vocational Decision Making Difficulty scale (VDMD; Holland, Gottfredson, & Nafziger, 1973), the CDS represents the earliest published attempt to assess level of and reasons for career indecision.

The CDS emerged from work beginning in a graduate seminar and evolved from an initial 14 items to its current 19-item format. Items one and two comprise the Certainty Scale and assess respondents' decidedness about, respectively, their career and academic major choices. Respondents rate themselves on these two items according to their levels of certainty and perceived comfort with and ability to implement their choices. Items 3-18 make up the Indecision Scale which assesses reasons for career indecision and correlates negatively with the Certainty Scale. Item 19 offers an open-ended response opportunity to clarify or elaborate on responses to the 18 preceding items. Osipow et al. (1976) designed the CDS primarily for high school and college students although, as Slaney (1988a) notes, it has been adapted successfully for use with graduate students, medical students, and non-traditional female college students. Extensive evidence exists for the reliability as well as the construct and concurrent validity of the measure (Slaney, 1988b). Counselors use the CDS to efficiently gauge clients' levels of decidedness and reasons for indecision, and to plan specific interventions based on item responses.

Many researchers have conducted factor-analytic studies of the CDS to determine whether its items scale different dimensions of indecision. If the CDS proved to measure different dimensions of indecision then counselors could use it to identify not only general indecision levels but also specific barriers to making career decisions. These factor analytic studies have fueled much debate about the utility of the CDS for this purpose. The dispute over the validity of the CDS as a multidimensional measure has produced a second generation of career certainty and career indecision measures.

Second-Generation Measures

Recent years have witnessed the emergence of a new age of career choice status measures. These measures differ significantly from earlier instruments in that researchers developed these later measures explicitly to assess multiple dimensions of career indecision. In so doing, these measures expanded Parsons' original model by operationally defining indecision as a multidimensional construct.

A revision of the Vocational Decision Scale, the Career Decision Profile (CDP; Jones, 1989) typifies measures designed specifically to scale different dimensions of career indecision and career choice status. Jones (1989) based the CDP on his and a colleague's earlier vocational decision status model. He showed in his initial validity study of the CDP that the vocational decision status model, consisting of three dimensions, "provides a clearer picture of career indecision than current unitary approaches" (p. 477). The CDP assesses respondents along the dimensions of (a) decidedness, or degree of certainty about choice, (b) comfort, or degree of contentment with decisional status, and (c) reasons, or basis for being decided or undecided.

The CDP Decidedness Scale contains two items on which respondents rate themselves using an 8-point scale. The first item contains content about having an occupational field in mind. The second item concerns having decided on an occupation to enter. Two additional items comprise the CDP Comfort Scale and contain content related to feeling at ease with or worried about career choice. Counselors can pair a client's scores on the scales of Decidedness and Comfort to profile a client's choice status as decided/comfortable, decided/uncomfortable, undecided/comfortable,undecided/uncomfortable. Four additional scales, each containing three items, assess respondents' reasons for their career uncertainty. These scales include (a) Self-Clarity, which concerns having knowledge about one's own interests, abilities, and so on, (b) Knowledge About Occupations and Training, which taps world-of-work knowledge, (c) Decisiveness, which measures ability to decide independently and resolutely, and (d) Career Choice Importance, which gauges feelings about the significance of work and making a career choice. Counselors can use these scales to identify specific barriers that prevent a client from reaching a career-decided state.

Summary and Conclusion

Since Parsons (1909) first classified people into career-decided and career-undecided groups, counseling researchers and practitioners have worked to formally assess career choice status. These efforts have yielded two generations of instruments useful for gauging clients' levels of andreasons for indecision as well as degrees of certainty about their career choices. Surveying clients in terms of their choice status continues to help researchers understand the complexity of career indecision and choice status. It also aids practitioners in planning appropriate career counseling interventions.

References

Holland, J. L., Gottfredson, G. D., & Nafziger, D. H. (1973). A diagnostic scheme for specifying vocational assistance (Rep. No. 164). Baltimore: The Johns Hopkins University, Center for Social Organization of Schools.

Jones, L. K. (1989). Measuring a three-dimensional construct of career indecision among college students: A revision of the Vocational Decision Scale—The Career Decision Profile. Journal of Counseling Psychology, 36, 477-486.

Osipow, S. H., Carney, C. G., Winer, J., Yanico, B., & Koschier, M. (1976). The Career Decision Scale (3rd Revision). Columbus, OH: Marathon Consulting and Press.

Parsons, F. (1909). Choosing a vocation. Boston: Houghton-Mifflin.

Savickas, M. L. (1992). "New directions in career assessment" in D. H. Montross & C. J. Shinkman (Eds.), Career development: Theory and practice (pp. 336-355). Springfield, IL: C. C. Thomas.

Slaney, R. B. (1988a). The assessment of career decision making. In W. B. Walsh, & S. H. Osipow (Eds.), Career decision making (33-76). Hillsdale, NJ: Lawrence Erlbaum Associates.

Slaney, R. B. (1988b). Review of the Career Decision Scale in J. T. Kapes and M. M. Mastie (eds.) A counselor's guide to career assessment instruments (2nd Edition) pp. 171-174. Alexandria, VA: National Career Development Association.

Williamson, E. G. (1937). Scholastic motivation and the choice of a vocation. School and Society, 46, 353-357.

Paul J. Hartung, Ph.D. is Assistant Professor of Behavioral Sciences, Northeastern Ohio Universities College of Medicine, Rootstown, Ohio.

ERIC Digest # EDO-CG-95-19

Gender Differences in Adolescent Career Exploration

Helen S. Farmer

Career exploration is a developmental stage identified by career development theorists (Super, 1990) and occurs typically during adolescence when boys and girls try out various work roles in part time work, volunteer work, or in school/community activities. Exploration tasks also include gaining an increasing awareness and understanding of the self and of abilities, interests, values, and needs. Jordaan (1963) indicated that exploration is the first of three substages leading to realistic career choice. Exploratory behavior follows the stage of tentative choice and is a time when people want to know as much as possible about themselves and about the world of work in order to make the best choice. This digest focuses on gender differences in the role of assessment in the exploration process. Career assessment texts, such as those of Walsh and Betz (1994) and Walsh and Osipow (1994), contain excellent chapters on gender bias in career assessment. In particular, the Gottfredson chapter in Walsh and Osipow provides extensive suggestions on how assessment may be used to stimulate career exploration that is gender fair.

Gender Differences in Career Exploration

Girls have been found typically to explore careers from a narrower set of career options than do boys. Gottfredson (1981) demonstrated how this occurs based on occupational sex role socialization. Girls and boys learn early which occupations are suitable for them and which ones are not. There have been concerted efforts on the part of educators, counselors, and the media to reduce occupational sex role stereotypes (Klein, 1985). Career education programs and classes in high school have attempted to reduce stereotyping in a variety of ways, including exposure to a wider variety of work environments, role models in nontraditional occupations, class discussion of issues related to occupational stereotyping and assessment of occupational interests in a gender neutral or sex fair way (Klein,1985). Increases in the participation of women in occupations nontraditional for them have occurred since the Educational Equity Act, and Equal Employment Legislation were passed in 1972. For example, women represented less than 1% of engineers in 1970, but, in 1990, women represented 17% of employed engineers (NSF, National Science Foundation, 1994).However, women are still seriously underrepresented in the higher paid, higher prestige, and better paying occupations, such as high level managers (i.e., CEO's), medical specialties involving surgery, the physical sciences and technical occupations (NSF, 1994). Occupational sex role socialization is still influencing the career exploration process for girls and boys.

Gender Differences in Career Interest Assessment

The most frequently used measures to aid in career exploration during adolescence are the career interest inventories. There are basically two kinds of interest measures, those based on empirical occupational scales such as the Strong Interest Inventory (SII), and those based on homogeneous scales such as the Self Directed Search (SDS) and the Kuder Occupational Interest Survey (KOIS). The former reflect the interests of persons currently in an occupation, that is , the status quo, and do not serve to stimulate exploratory behavior as well as the homogeneous scaled inventories, which provide, for each interest, a measure of how similar a person's interests are to a set of items that all assess that interest (for example, artistic interest). The concept of "exploration validity" based on the extent to which an interest inventory stimulates the person to explore career options that might otherwise not be explored is relevant to the gender issues discussed in this digest. Interest inventories were criticized in the 1970s because they typically used sexist language and items that were biased toward men and yielded scores that rarely encouraged girls to explore occupations nontraditional for their gender (Diamond,1975). The National Institute of Education (NIE) issued guidelines for reducing sex bias in interest measurement (Diamond, 1975) and these guidelines were effective in stimulating the publishers of the most frequently used career interest measures to revise their instruments to make them more sex fair (i.e., Strong Interest Inventory(SII), Harmon, Hansen, Borgen, & Hammer,1994; Kuder Occupational Interest Survey (KOIS), Kuder & Zytowski,1991; and The Self Directed Search(SDS), Holland, Fritzsche, & Powell,1994). Sex bias was defined in the NIE Guidelines (Diamond, 1975) as "any factor that might influence a person to limit—or might cause others to limit—his or her consideration of a career solely on the basis of gender." These guidelines further suggested that administration of an interest inventory be accompanied by an orientation dealing with possible influences from the environment, culture, early socialization, traditional sex role expectations of society, home-versus-career conflict, and the experiences typical of women and men as members of various ethnic and social class groups on men's and women's scores. Such orientation should encourage respondents to examine stereotypic "sets" toward activities and occupations and should help respondents to see that there is virtually no activity or occupation that is exclusively male or female (Diamond,1975, pp. xxvi-xxvii). Interest inventories that extend exploration of occupations beyond those the client has already considered into fields not typical for their gender would be viewed as responsive to the NIE guidelines. Which interest inventories in 1994 best meet this exploratory validity criterion?

During the period from the early 1970s to the mid 1980s most interest measures met the criteria set down by the NIE guidelines to eliminate sexist language, to use the same form of the test for both sexes; to provide scores on all occupational scales for both sexes with an explanation of which norms were used to develop the scale, and to use items that equally reflected the experiences/activities familiar to both sexes.

Not surprisingly, perhaps, career interest inventories such as the Self Directed Search (Holland, et al. 1994) still obtain significantly higher scores for women on Social scales (i.e., those related to people and service oriented occupations) and significantly higher scores for men on Realistic scales (i.e. those related to technical, skilled trades, engineering occupations). Hansen, Collins, Swanson, and Fouad (1993) assessed sex differences in Holland's hexagon ordering of career interests as measured by the SII and found that the distance between interest types was significantly different for men and women when samples were matched for occupation and level. These authors found that women's scores on Investigative and Realistic scales were highly correlated and that the structure of Holland's Hexagon was significantly different for men and women. The SII (Harmon et al., 1994) Manual suggests the use of this inventory to facilitate career exploration for the non-college bound youth, but not for the college bound. Since evidence of gender differences continue to be found for career interest measures it seems imperative to revive the NIE guidelines orienting women clients to the effects of their socialization on their scores. In the latest version of the SDS the Assessment Booklet gives the following advice to users **after** they have obtained their SDS scores: "Remember that results are affected by many factors in your background. For example, because society encourages men and women to aspire to different vocations women receive more Social, Artistic and Conventional codes than men, while men receive more Investigative, Realistic and Enterprising codes. Yet we know that almost all jobs can be successfully performed by members of either sex. If your codes differ from your Occupational Daydream codes keep these influences in mind. You may decide to stick with your Daydreams" (Holland,1994, p. 12). It would be interesting to know what kind of SDS scores a person might obtain if they received this message **before** taking the inventory, consistent with NIE Guidelines.

Summary

The NIE guidelines for reducing sex bias in interest measurement (Diamond, 1975) were followed to a large extent by both interest measurement test developers and publishers in the decade following their publication. The concept of "exploration validity," the extent to which an interest inventory stimulates the person to explore career options that might otherwise not be explored has been widely adopted. However, the continuing evidence that gender differences exist in career interest measurement strongly suggests that such assessment is accompanied with counseling. The NIE guidelines (Diamond, 1975) suggesting that exploration during adolescence should expand beyond the social learning experiences of an individual, and beyond their expressed interests, to include exposure to other career options that sex equity legislation has opened up to women should be followed if career exploration is to become more gender fair.

References

Diamond, E. (1975). *Issues of sex bias and sex fairness in career interest measurement*, Washington, DC: U.S. Department of Education.

Diamond, E. & Zytowski, D. (1991)._*Manual for the Kuder Occupational Interest Survey, DD*. Chicago, IL: Science Research Associates.

Gottfredson, L. (1981). Circumscription and compromise: A developmental theory of occupational aspirations (Monograph) *Journal of Counseling Psychology, 28*, 545-579.

Hansen, J. I., Collins, R., Swanson, J., & Fouad, N. (1993). Gender differences in the structure of interests. *Journal of Vocational Behavior, 42*, 200-211.

Harmon, L., Hansen, J.I, Borgen, F., & Hammer, A. (1994)._*Strong Interest Inventory: Applications and Technical Manual*, Palo Alto, CA: Consulting Psychologists Press.

Holland, J., Fritzsche, B. & Powell, A. (1994) *Technical Manual for the Self-Directed Search*, Lutz, FL: Psychological Assessment Resources.

Jordaan, J. (1963). Exploratory behavior: The formation of self and occupational concepts. In D. Super, R. Starishevsky, N. Matlin, & J. Jordaan, (pp. 42-78) New York, NY: The College Board.

Klein, S. (1985). *Handbook for achieving sex equity through education.* Baltimore, MD: Johns Hopkins University Press.

National Science Foundation (1994). *Request for Proposals.* Washington, DC: U.S.Government Printing Office.

Super (1990). A life-span, life-space approach to career development. In D. Brown , L. Brooks, & Associates. *Career choice and development* (2nd ed, pp 197-261). San Francisco, CA: Jossey-Bass.

Walsh, B. & Betz, N. (1994). *Tests and assessment.* Englewood Cliffs, NJ: Prentice-Hall.

Walsh, B. & Osipow, S. (1994). *Advances in vocational psychology:* Volume 1: *The assessment of interests.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Helen S. Farmer, Ph.D., is a Professor, Department of Educational Psychology, Counseling Psychology Division, University of Illinois at Urbana-Champaign.

ERIC Digest # EDO-CG-95-20

Assessing Employability Skills

Thomas H. Saterfiel and Joyce R. McLarty

Employability skills refers to those skills required to acquire and retain a job. In the past, employability skills were considered to be primarily of a vocational or job-specific nature; they were not thought to include the academic skills most commonly taught in the schools. Current thinking, however, has broadened the definition of employability skills to include not only many foundational academic skills, but also a variety of attitudes and habits.

In fact, in recent usage, employability skills is often used to describe the preparation or foundational skills upon which a person must build job-specific skills (i.e., those that are unique to specific jobs). Among these foundational skills are those which relate to communication, personal and interpersonal relationships, problem solving, and management of organizational processes (Lankard, 1990). Employability skills in this sense are valued because they apply to many jobs and so can support common preparation to meet the needs of many different occupations.

The concept of employability skills originated with educators, primarily those working on programs specifically designed to facilitate employment (e.g., vocational rehabilitation, Job Training Partnership Act). Employers, although the primary determiners of the skills that will actually enable an individual to acquire and retain a job, have traditionally focused on job-specific skills (e.g., those needed to spot weld or prepare a sales report). Assessments for employment, where used, most frequently have consisted of general ability and personality tests supplemented by job-specific assessments (e.g., work samples).

In recent years, that picture has changed dramatically with ever-growing numbers of employers assessing foundational skills, primarily in reading and mathematics, prior to hiring (Greenberg, Canzoneri, and Straker, 1994). This is probably due to the joint effects of an increasing demand for these skills on the job and employer dissatisfaction with the levels of those skills demonstrated by applicants. Even today, however, educators show greater interest in employability skills assessment than do employers. This is possibly due to employer concerns about the legal implications of any assessment that might have an adverse impact (a detrimental effect on hiring rates) on gender or ethnic minority groups (Uniform Guidelines, 1978).

Much of the current impetus to teach and assess employability skills results from concerns about this country's ability to compete in the world economy. Seminal work by Carnevale (Carnevale, Gainer, and Meltzer, 1990) was followed by efforts by both public and private agencies to address the strongly felt need to improve the work-related skills of those entering the workforce. The work begun by the Department of Labor and its Secretary's Commission on Attaining Necessary Skills (SCANS) is continuing, with plans to validate the skills they identified (U.S. Department of Labor, SCANS, 1992). Development of assessments for these skills will follow this effort.

American College Testing's Center for Education and Work, through its Work Keys System, has developed large-scale assessments for seven employability skill areas: Reading for Information, Applied Mathematics, Listening,Writing, Locating Information, Applied Technology, and Teamwork. Assessments for additional skill areas are currently in development (American College Testing, 1994). The state of Ohio combined its job specific Ohio Compentency Assessment Program (OCAP) tests with the Work Keys assessments for a comprehensive assessment of foundational and specialized skills. The state of Tennessee is involving its high school seniors in the Work Keys System to help it meet the employability skills needs of all its students.

Other notable efforts include the C3 project in Fort Worth, Texas (Fort Worth Independent School District, 1992) and the portfolio development and evaluation undertaken by the state of Michigan (Michigan Occupational Information Coordinating Committee, 1992). These projects are distinguished by extensive use of business input for development and implementation. Although neither of these projects currently offers assessments for use by outside agencies, both are sources of valuable information on the development of employability skills.

Of the many other efforts to provide employability skills assessments, the largest number focus on the basic literacy level, as did the earliest work on employability skills.Educational Testing Service, building on the work of the National Adult Literacy Study funded by the U.S. Department of Education, publishes tests measuring lower-level reading, mathematics, and document literacy. Additionally, tests once used only for assessing lower-level adult skills for academic purposes have now also been pressed into service to meet the growing demand for employability skills assessment (e.g., TABE, CASAS).

When selecting an approach for assessing employability skills, several criteria must be kept in mind.First, the validity of an employability skills assessment rests on job analysis: a clear and validated relationship should exist between the assessment and the skills required for one or more jobs. This relationship should be based on a systematic analysis of the skills and skill levels required for the job(s) in question. It is not sufficient to observe, for example, that "reading" is required for the job; one must know which tasks require reading and the type and level of reading skill needed. The assessment must clearly mirror the nature of the skill required, and the score attained on it must accurately reflect the examinee's level of that skill.

Second, the skill assessed should be teachable. Assessment of "intrinsic abilities" is valuable both for employers attempting to predict future job performance and for counselors working with students to identify jobs suited to their interests, values, and self-concepts. However, the essence of employability skills is preparation for the job, so the focus of employability skills assessments should be directed to those aspects of the relevant skills that can be taught. Since not all employability skills can be neatly packaged in the traditional academic disciplines, educators must make special efforts to ensure that they teach all the needed employability skills.

The degree to which preparation for the workforce (i.e., employability skills development) and preparation for postsecondary education are congruous has been under considerable discussion. It is too early to determine whether integrated preparation for both provides as good a preparation for each as separate programs or, if not, at what point in a student's career separate programs should begin. Institutions using separate programs for preparation generally begin that differentiation at grade 10 or 11.

Finally, each assessment must be evaluated in the context of its purpose. If employers are going to use the scores to make personnel decisions, the employability skills assessment must meet strict reliability and validity standards, sufficient to provide a sound legal defense. This requires painstaking attention to the psychometric quality of the instrument, to the standardization of the administration, and to the accuracy of the scoring. However, if the purpose of the assessment is to guide instruction, relevant psychometric criteria are more relaxed. The advantage of assessments which employers may use for personnel decisions is that the results are of immediate use to the examinees in making the transition to the workforce. The advantage of assessments used only for low stakes purposes is that they may be constructed with greater emphasis on providing instructionally relevant experiences to students. It is also important to recognize that assessment instruments are needed to support the information needs both of school-age students as they enter the workforce and of adults making transitions into, or within, the workforce at later stages in their lives.

References

American College Testing. (1994). The Work Keys system. [Brochure]. (Available from ACT Work Keys ClServices, P.O. Box 168, Iowa City, IA 52243-0168)

Carnevale, A. P., Gainer, L. E., & Meltzer, A. S. (1990). Workplace basics: The essential skills employers want. San Francisco, CA: Jossey-Bass Publishers.

Fort Worth Independent School District (1992). Making education work: Vital Link. [Brochure]. (Available from Fort Worth Independent School District, 3210 West Lancaster, Fort Worth, TX 76107)

Greenberg, E. R., Canzoneri, C., & Straker, T. (1994). 1994 AMA survey on basic skills testing and training.(Available from the American Management Association, 135 W. 50th Street, New York, NY 10020)

Lankard, B. A. (1990). Employability—the fifth basic skill. ERIC Digest No. 104. Columbus: Center on Education and Training for Employment. The Ohio State University. (ERIC No. EDO-CE-90-104)

Michigan Occupational Information Coordinating Committee. (1992, September). How do I get from here to there: A guide to work-based learning. (Available from MOICC, c/o Michigan Department of Labor, Box 30015, Lansing, MI 48909)(ERIC Document Reproduction Service No. ED 361 621)

Uniform guidelines in employee selection procedures. (1978). Federal Register, 43, 38290-38315.

U.S. Department of Labor, Secretary's Commission on Achieving Necessary Skills. (1992). Learning a living:A blueprint for high performance. Washington, D.C.: Government Printing Office.

ERIC Digest # EDO-CG-95-21

Assessing Career Development With Portfolios

Juliette N. Lester and Nancy S. Perry

The assessment of career development is a relatively new concept. In general, ideas of appropriate methods for assessing student achievement and mastery of any set of competencies are shifting. Criterion-referenced tests, which measure performance relative to a specified set of standards or tasks, are gaining favor, for example, over norm referenced tests, which measure how an examinee performs in relation to others. At the same time, support for internal accountability —that is, determining what is worth knowing and assuring that students know it— is increasing. One response to this has been an increased use of portfolios that provide a medium for assessing student work and invite them to become responsible partners in documenting their learning. Through portfolios, students compose a portrait of themselves as able learners, selecting and presenting evidence that they have met the learning standards for individual classes and for broader learning tasks (Wolf, LeMahieu & Eresh, 1992). A student portfolio may be described as "a purposeful collection of student work that tells the story of the student's efforts, progress, or achievement in a given area. This collection must include student participation in selection of portfolio content; the guidelines for selections; the criteria for judging merit; and evidence of student self-reflection" (Arter and Spandel, 1992, p. 36).

As career development becomes an increasingly important component of educational systems, the issues of measurement and accountability are raised. This digest focuses on the use of portfolios in assessing career development.

Career Development Goals

In today's workplace, employment security is becoming "employability security" (Kanter, 1991, p.9) —the knowledge that one has the competencies demanded in a global economy and the ability to expand and adjust those competencies as requirements change. The challenge of preparing our young people for this new workplace has generated legislative efforts to stimulate educational reform directed at creating "world class" education and a comprehensive system for helping American youth make a smooth transition from high school to productive, skilled employment and further learning. The Goals 2000: Educate America Act establishes eight national education goals and two national councils —one to stimulate the development of voluntary academic standards and the other to identify essential occupational skills. The School to Work Opportunities Act of 1994 is a strategy to implement the purpose of the Goals 2000: Educate America Act, that is, helping all Americans to reach internationally competitive standards through educational reform.

Career development is a major component of the School to Work Opportunities Act (STWOA). Career guidance and counseling, which are interventions in the career development process, are recognized as essential in helping students to choose their career (educational) pathway. Section 102 of the STWOA states that "The school-based component of a School-to Work Opportunities program shall include ... career awareness and career exploration and counseling (beginning at the earliest possible age, but not later than the 7th grade) in order to help students who may be interested to identify, and select or reconsider, their interests, goals, and career majors, including those options that may not be traditional for their gender, race or ethnicity." The Act also provides grants to states to plan for and implement school-to-work opportunities systems.

Renewed interest in career development has led to an equal demand for accountability. This prompts several questions. What do we want our students to know and be able to do as a result of a career development process, and how will we know that they have achieved it? This legislation has placed the onus on school systems to provide the programs to help students make informed career decisions, and to provide opportunities for students to take responsibility for their career development. How will they know they have achieved these outcomes?

Two major endeavors can help schools to meet the double need of accountability and assessment. First, state and professional associations, as well as national leaders, practitioners, and career development experts, collaborated to develop the National Career Development Guidelines (NOICC, 1989). The National Career Development Guidelines offer a comprehensive, competency-based approach to career development that states, educational institutions, and other organizations can use in developing effective career guidance programs. The Guidelines offer the processes, content, and structure for such programs. More importantly, they provide the standards or competencies for career development at four different levels—elementary, middle/junior high, high school, and postsecondary/adult. The competencies fall within three areas of career development—self-knowledge, educational and occupational exploration, and career planning. The Guidelines, already being used in over 40 states as standards or as the basis for establishing career development standards, provide nationally validated competencies that can be used in assessment.

The second significant effort has been the work of the Secretary's Commission on Achieving Necessary Skills (SCANS). In the Commission report, What Work Requires of Schools (U. S. Department of Labor, 1991), five areas of competencies based on a three-part foundation are delineated. Of the 36 specific skills or qualities noted, over half are commonly included in a career guidance program. This report validates the integration of career guidance and counseling into educational programs and supplies a complementary set of standards by which a career development process can be measured.

Assessment Through Portfolios

The essential criteria for measuring the accountability of a career guidance program are available. Since self-assessment and reflection are important to developing personal responsibility in career decision-making, a portfolio that sets standards and also allows for reflection emerges as the instrument of choice. Until now, most efforts to document career development have been through career planners. Career planners are usually the end product of a career development process and, as such, are appropriate for secondary education or higher but not for the student at the awareness or exploratory stages. They also do not typically provide for the self-reflection essential to an individual's ownership of the process.

Get a Life: Your Personal Planning Portfolio (ASCA, 1993), designed through collaboration between the American School Counselor Association and the National Occupational Information Coordinating Committee, is one instrument that sets standards and allows for self-reflection. The portfolio is divided into four sections —self-knowledge, life roles, educational development, and career exploration and planning. Each section contains competency files and personal files. The National Career Development Guidelines for the middle and high school levels are used as competencies for both program and individual assessment. Program planners can analyze the comprehensiveness of their programs by evaluating their activities in relation to the expected student outcomes contained in the Guidelines. Individuals can determine if they have met the career development competencies through the programs offered. Within the competency file, a sign-off ascertains the strategies and the date on which each competency was addressed. In some schools, students make the decision whether, in fact, the activity or strategy presented did help them to master the competency. The personal files are a set of guiding questions that help students to reflect on their learning. The portfolio is an organizational tool that allows the owners to collect information about themselves to use in making personal, educational and career decisions. At the same time, the students are introduced to the idea that the process is lifelong, and that they must become "career negotiators" (Bailyn, 1992), taking responsibility for their own development.

Summary and Conclusion

Recent efforts to improzpport student learning, and assess both the program and the individual to assure that the expected outcomes are being achieved. The portfolio provides the format for the process and documentation of career development while giving individuals and programs standards for assessment.

References

American School Counselor Association. (1993). *Get a life: Your personal planning portfolio*. Alexandria, VA: Author.

Arter, J. & Spandel, V. (1992). NCME Instructional Module: Using portfolios of student work in instruction and assessment. *Educational Measurement: Issues and Practice, 11*(1), 36-44.

Bailyn, L. (1992). "Changing the conditions of work." *In career development: theory and practice*. D. Montross and C. Shinkman, (Eds.). Springfield, IL: Charles C. Thomas.

Kanter, R. (1991). Globalism/localism. *Harvard business review, 69*(2), 9-10.

National Occupational Information Coordination Committee (1989). *National career development guidelines*. Washington, DC: Author. (ERIC Document ReproductionService Nos. ED317874-ED317880)

U.S. Department of Labor. (1991). *What work requires of schools*. Washington, DC: Author.

Wolf, D., LeMahieu, P., & Eresh, J. (1992, May). Good measure: Assessment as a tool for education reform. *Educational Leadership, 49*(8), 8-13.

Juliette N. Lester is the Executive Director of the National Occupational Information Coordinating Committee (NOICC).

Nancy S. Perry, Guidance Consultant for the Maine Department of Education, is currently serving as Career Development Specialist with NOICC.

ERIC Digest #EDO-CG-95-22

Ethics in Assessment

Cynthia B. Schmeiser

Every profession has distinct ethical obligations to the public. These obligations include professional competency, integrity, honesty, confidentiality, objectivity, public safety, and fairness, all of which are intended to preserve and safeguard public confidence. Unfortunately, all too often we hear reports in the media of moral dilemmas and unethical behavior by professionals. These reports naturally receive considerable attention from the public, whose confidence in the profession is undermined with each report.

Those who are involved with assessment are unfortunately not immune to unethical practices. Abuses in preparing students to take tests as well as in the use and interpretation of test results have been widely publicized. Misuses of test data in high stakes decisions, such as scholarship awards, retention/promotion decisions, and accountability decisions, have been reported all too frequently. Even claims made in advertisements about the success rates of test coaching courses have raised questions about truth in advertising. Given these and other occurrences of unethical behavior associated with assessment, the purpose of this digest is to examine the available standards of ethical practice in assessment and the issues associated with implementation of these standards.

Existing Ethical Standards

Concerns about ethical practices in assessment are not new. As early as 1972, the National Council on Measurement in Education (NCME), the Association for Measurement and Evaluation in Guidance (AMEG), and the American Association for Counseling and Development (AACD is now known as theAmerican Counseling Association) developed a position paper on the responsible use of tests that was intended to ensure that tests are given, and examinees are treated, fairly and wisely (AMEG, 1972). Later in the 1970s, AACD developed a statement on the responsibilities of the users of standardized tests, a document that was revised in 1989 (AACD, 1989). Both of these early documents recognized the need to positively influence the practices of those who use tests in ways that promote responsible use.

These statements have been followed by the development of ethical standards by a number of other organizations having an interest, or directly involved, in assessment. These standards address assessment practices and related issues for various professionals: psychologists (American Psychological Association, 1992); counselors (American Association for Counseling and Development, 1988, 1989); educational researchers (American Educational Research Association, 1992); teachers (American Federation of Teachers, National Council on Measurement in Education, National Education Association, 1990); measurement specialists (American Educational Research Association, American Psychological Association, National Council on Measurement in Education 1985; Joint Committee on Testing Practices, 1988); educational evaluators (Joint Committee on Standards for Educational Evaluation, 1988); evaluators of educational programs (Joint Committee on Standards for Educational Evaluation, 1994); college admission counselors (National Association of College Admission Counselors, 1988); and others. The National Council on Measurement in Education is considering the adoption of a Code of Professional Responsibilities in Educational Measurement in the fall of 1994.

All of these codes vary widely in their scope: some include technical standards that the professionals should meet in their practice, but all of them include some statements about ethical responsibilities that are intended to guide the behavior of professionals as they use assessments in their practice. The codes focusing exclusively on ethics that have been adopted by professions are intended to clarify the expectations of professional conduct in various situations encountered in practice and to affirm that the profession intends and expects its members to recognize the ethical dimensions of their practice. The fact that all of these standards exist is evidence that these organizations are seriously concerned about and committed to promoting high technical standards for assessment instruments and high ethical standards for individual behavior as professionals work with assessments.

In recent years, there have been increasing discussions in the professions about how to make sure that proper ethical conduct is not only advocated as an ideal but also practiced. Yet, even once a code of ethics has been adopted, each organization has had to struggle with issues of both enforcement and education.

To Enforce or Not To Enforce?

Whether a code of ethics will be enforced and how it will be enforced has been a dilemma for most organizations. Even with the codes cited earlier, there is a great deal of variability in the approaches taken by the adopting organizations to enforce the codes. There appears to be at least four general approaches to enforcement.

First, some organizations have no formal enforcement of their codes; the standards are designed to increase the awareness of their members as to what constitutes ethical practice and to serve as an affirmation of exemplary conduct. Organizations like AERA and NCME have no formal enforcement mechanism, typically have no sanctions attached to membership in the organization, and membership is not tied to a credential in any way.

Second, some organizations enforce their codes of ethics at the local level. The national organizations delegate enforcement to affiliated state societies that have adopted the national code in whole or in part as their state society's code of ethics. This type of enforcement is used, for example, by the legal profession in that the American Bar Association's ethical codes serve as model legislation for state bars to use in creating and enforcing their own codes.

Third, some organizations enforce their codes at the national level. The ways in which enforcement is handled at the national level vary significantly. Organizations like the American Counseling Association and the American Psychological Association have established special divisions or committees as enforcement arms. Other organizations have established trial boards that adjudicate disciplinary charges and impose discipline; in other organizations, local chapters refer cases to the national ethics committee for adjudication and possible discipline.

The fourth model involves enforcement at both the national and local levels. For instance, the American Medical Association might take disciplinary action against a member when the state medical association to which the physician belongs requests or consents to such action. At this time, however, there does not appear to be an assessment-related organization that uses this type of enforcement.

The approach taken by a professional organization to enforce its code of ethics is usually directly related to the purpose of the code and the requirements for practice. If membership in the organization is voluntary, it is difficult to establish a formal means of discipline and enforcement. Certainly, membership in such an organization could be revoked, but it would not prevent the member from practicing. By contrast, when membership in the professional organization is tied to a credential or a designation of some type, then establishing a formal means of discipline and enforcement (such as formal/informal reprimands, revocation of designation, or expulsion from the profession) is easier to establish and implement.

To Educate

Nearly all organizations that have adopted a code of ethical assessment practices engage in educational activities that are intended to promote a greater understanding of what constitutes ethical assessment practice. Educational activities are particularly important since a code of ethics is not a set of givens, but rather a frame of reference for the evaluation of the appropriateness of behavior. Case studies can serve as particularly effective illustrations of how ethical issues may be analyzed and how judgment may be used to evaluate behavior. Other effective educational approaches include open forums for discussions of ethical issues, disseminating realistic problems that involve judgments about appropriateness of behavior, and group learning activities that pose ethical dilemmas that are analyzed and evaluated by groups of professionals. Regardless of the approach taken, dissemination of the codes supported by real-life examples of ethical dilemmas are effective ways of promoting an understanding of ethical assessment practice.

Summary

The level of enforcement that each organization imposes is directly tied to the character of membership in the organization, whether it is voluntary or tied to a credential or designation. Clearly, the more stringent the requirements are for membership in an organization, the easier it is for that organization to establish a formal means of discipline and enforcement.

Educating others to understand and to engage in ethical practices is a critical goal. Illustrations of good and bad practice within realistic assessment contexts and discussions of ethical dilemmas are excellent ways of promoting ethically responsible practice in assessment.

References

American Association for Counseling and Development (now American Counseling Association). (1988). Ethical standards of the American Counseling Association. Alexandria, VA: Author.

American Association for Counseling and Development (now American Counseling Association) and Association for Measurement and Evaluation in Counseling and Development (now Association for Assessment in Counseling). (1989). Responsibilities of users of standardized tests: RUST statement revised. Alexandria, VA: Author.

American Educational Research Association. (1992). Ethical standards of the American Educational Research Association. Washington, DC: Author.

American Educational Research Association, American Psychological Association, National Council of Measurement in Education. (1985). Standards for educational and psychological testing. Washington, DC: APA.

American Federation of Teachers, National Council on Measurement in Education and National Education Association. (1990). Standards for teacher competence in educational assessment of students. Washington, DC: Author.

American Psychological Association. (1992). *Ethical principles of psychologists and code of conduct.* Washington, DC: Author.

Joint Committee on Standards for Educational Evaluation. (1988). The personnel evaluation standards: How to assess systems for evaluating educators. Newberry Park, CA: Sage.

Joint Committee on Standards for Educational Evaluation (1994). The program evaluation standards: How to assess evaluations of educational programs. Thousand Oaks, CA: Sage.

National Association of College Admission Counselors. (1988). Statement of principles of good practice. Alexandria, VA: Author.

The responsible use of tests: A position paper of AMEG, APGA and NCME (1972). *Measurement and Evaluation in Guidance, 4*(2), 385-388.

Cynthia Board Schmeiser, Ph.D., is Vice President of the Development Division of American College Testing, 2201 N. Dodge Street, Iowa City, IA 52243, and is currently chair of the NCME Code of Ethics Development Committee.

ERIC Digest # EDO-CG-95-23

Multicultural Assessment

William E. Sedlacek and Sue H. Kim

Assessment includes the use of various techniques to make an evaluation; *multicultural assessment* refers to the cultural context in which the assessment is conducted, namely one in which people of differing cultures interact. One can argue that all assessments are conducted and interpreted within some cultural context, but only recently have the cultural assumptions underlying such assessments been acknowledged (Sue & Sue, 1990). The fields of counseling and therapy traditionally have relied heavily upon the use of assessment techniques to gather information about clients in order to indicate appropriate directions for treatment. Measures to assess personality, cognitive abilities, interests, and other psychological constructs have been utilized in a variety of different counseling and education settings. Although many of the measures most widely used have established reliability and validity only within White racial samples, these measures often are used inappropriately and unethically with populations from different cultures.

This digest identifies four common misuses of assessments in multicultural contexts, describes some of the ways in which multicultural assessments can be improved, and suggests topics for future research in the area of multicultural assessment.

Common Misuses of Assessments in Multicultural Contexts

1. Assuming that labeling something solves the problem. Sedlacek (in press, a) has called this the "Quest for the Golden Label" problem. Using new terms (e.g., multicultural, diversity) does not mean we are doing anything operationally different with our measures. Westbrook and Sedlacek (1991) found that although labels for nontraditional populations had changed over forty years, the groups being discussed were still those without power who were being discriminated against in the system.

2. Using measures normed on White populations to assess non-White people. Sedlacek (in press, a) discussed what he called the "Three Musketeers" problem, namely that developing a single measure with equal validity for all is often the goal of test developers. However, if different people have different cultural and racial experiences and present their abilities differently, it is unlikely that a single measure could be developed that would work equally well for all.

3. Ignoring the cultural assumptions that go into the creation of assessment devices. Helms (1992) argued that cognitive ability measures are commonly developed from an unacknowledged Eurocentric perspective. Until there is more thought given to the context in which tests are developed, work comparing different racial and cultural groups using those measures will be spurious.

4. Not considering the implications of the use of measures with clients from various racial and cultural groups. Professionals may not be adequately trained in determining which measures are appropriate to use with particular clients or groups. Sedlacek (in press, a) has called this the "I'm OK, you're not" problem in which very few professionals receive adequate training in both instrument development and an appreciation of multicultural issues.

Suggestions for Improving Multicultural Assessments

1. Concentrate on empirical and operational definitions of groups, not just labels. Sedlacek (in press, b) has suggested that if members of a group receive prejudice and present their abilities in nontraditional ways, they can be considered "multicultural." He suggested the use of measures of racial attitudes and noncognitive variables in making this determination.

2. Identify measures specifically designed for multicultural groups. Sabnani and Ponterotto (1992) provided a critique of "racial/ethnic minority-specific" instruments and made recommendations for their use in different assessment contexts. Prediger (1993), in a compilation of multicultural assessment standards for counselors developed for the American Counseling Association, recommended that a determination be made that the assessment instrument was designed for use with a particular population before it is used.

3. Encourage the consideration of cultural factors in the earliest conceptual stages of instrument development. Helms (1992) called this a "culturalist perspective" in assessment. Sedlacek (in press, a) noted a lack of developmental multicultural thinking as new instruments are developed. Multicultural groups are usually "throw ins" after the fact to see how their test results compare with those of the population on which the test was normed. He called this the "Horizontal Research" problem in developing assessment measures.

4. Increase opportunities for an exchange of information between those with quantitative training in instrument development and those with an interest and expertise in multicultural issues. Currently there is

little overlap in these two groups. Helms (1992) felt it was important not to assume that there are enough professionals of color to do this work. Many individuals from majority racial and cultural groups will need to develop such measures as well. Conventions, workshops, coauthored articles, and curricular reform in graduate programs are but a few examples of what could be done.

Topics for Future Research on Multicultural Assessment

Research on the validity and reliability of measures for specific multicultural groups is needed (Helms, 1992; Sabnani & Ponterotto, 1992). This includes studies of attributes that may be more important for multicultural groups than for others. Noncognitive variables, such as handling racism or having support of a cultural or racial group, have been shown to be particularly useful for members of nontraditional groups and should be studied further. Additional research on the utility of defining nontraditional groups broadly to include diversity based on age, physical disability, sexual orientation, etc. (Sedlacek, in press, a), or to concentrate on the major racial and cultural groups, e.g., African Americans, American Indians, Asian Americans, and Hispanics (Sue, Arredondo, & McDavis, 1992) should be conducted.

Summary

More valid assessments for multicultural populations would help counseling professionals better serve their clients and improve the lives of many people whose backgrounds and experiences may differ from those of White clients. Four common misuses of assessments in multicultural contexts were presented here, as were ways to counteract those misuses. Concentrating on empirical and operational definitions of multicultural groups rather than relabeling was the first suggestion discussed. Using measures specifically designed for multicultural groups was recommended as the best solution to the problem of using instruments normed on White populations. Developing new measures from a "culturalist perspective" was the recommended way to counter a lack of multicultural thinking in instrument development. Creating more opportunities to bring together those with training in instrument development and those with multicultural interests was seen as a way to improve the quality of multicultural assessments by professionals.

References

Helms, J.E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? American Psychologist, 47(9), 1083-1101.

Prediger, D.J. (1993). Multicultural assessment standards: A compilation for counselors. Alexandria, VA: American Counseling Association.

Sabnani, H.B., & Ponterotto, J.G. (1992). Racial/ethnic minority-specific instrumentation in counseling research: A review, critique, and recommendations. Measurement and Evaluation in Counseling and Development, 24(4), 161-187.

Sedlacek, W.E. (in press, a). Advancing diversity through assessment. Journal of Counseling and Development.

Sedlacek, W.E. (in press, b). An empirical method of determining nontraditional group status. Measurement and Evaluation in Counseling and Development.

Sue, D.W., Arredondo, P., & McDavis, R.J. (1992). Multicultural counseling competencies and standards: A call to the profession. Journal of Counseling and Development, 70(4), 477-486.

Sue, D.W., & Sue, D. (1990). Counseling the culturally different: Theory and practice. New York: Wiley.

Westbrook, F.D., & Sedlacek, W.E. (1991). Forty years of using labels to communicate about nontraditional students: Does it help or hurt? Journal of Counseling and Development, 70(1), 20-28.

William E. Sedlacek, Ph.D., is Professor of Education and Assistant Director, Counseling Center, University of Maryland at College Park.

Sue H. Kim, M.Ed., is Research Assistant, Counseling Center, University of Maryland at College Park.

ERIC Digest # EDO0CG-95-24

Fairness in Performance Assessment

Tony C. M. Lam

Performance assessment is the type of educational assessment in which judgments are made about student knowledge and skills based on observation of student behavior or inspection of student products (see the digest by Stiggins in this series). Education reformers have hailed policy that pushes performance assessment as manna (miraculous food) from above, feeding teachers and students "wandering in a desert of mediocrity" (Madaus, 1993, p. 10). They claim that by replacing selection response tests such as multiple-choice tests with performance assessment, our schools will improve, and all ailments in student assessment, including the affliction of unfairness, will be cured. Unfortunately, although the pedagogical advantages of performance assessment in supporting instruction that focuses on higher-order thinking skills are obvious, research has consistently indicated unresolved logistic and psychometric problems, especially with score generalizability (Linn, 1993). In addition, there is no evidence suggesting that assessment bias vanishes with performance assessment (Linn, Baker, & Dunbar, 1991).

Bias & Fairness

Consonant with the unified conceptualization of validity (Messick, 1989), assessment bias is regarded as differential construct validity that is addressed by the question: To what extent is the assessment task measuring the same construct and hence has similar meaning for different populations? The presence of bias invalidates score inferences about target constructs because of irrelevant, non-target constructs that affect student performance differently across groups. These irrelevant constructs are related to characteristics such as gender, ethnicity, race, linguistic background, socioeconomic status (SES), or handicapping conditions that define the groups. For example, ability to read and understand written problems is a biasing factor in measuring mathematics skills because it is irrelevant to mathematics skills and it affects Limited English Proficient (LEP) and native English speaking students' performance differently on a mathematics test.

Assessment for its intended purpose is unfair if (1) students are not provided with equal opportunity to demonstrate what they know (e.g., some students were not adequately prepared to perform a type of assessment task) and thus the assessments are biased; (2) these biased assessments are used to judge student capabilities and needs; and (3) these distorted views of the students are used to make educational decisions that ultimately lead to limitations of educational opportunities for them. Despite a common definition of assessment fairness in reference to assessment bias, the approach and methods used to assure fairness are nevertheless determined by one's choice of either one of two antithetical views of fairness: equality and equity.

Equality

The equality argument for fairness in assessment advocates assessing all students in a standardized manner using identical assessment method and content, and the same administration, scoring, and interpretation procedures. With this approach to assuring fairness, if different groups of test takers differ on some irrelevant knowledge or skills that can affect assessment performance, bias will exist.

Traditional tests with selection response items have been criticized as unfair to minority students because these students typically perform less well on this type of test than majority students. However, no evidence is yet available to substantiate the claim that performance assessment can in fact diminish differential performance between groups (Linn et. al., 1991). Although the use of performance assessment can eliminate some sources of bias, such as testwiseness in selecting answers that are associated with traditional tests, it fails to eliminate others, such as language proficiency, prior knowledge and experience, and it introduces new potential sources of bias: (1) ability to handle complex problems and tasks that demand higher order thinking skills (Baker & O'Neil, 1993); (2) metacognitive skills in conducting self-evaluation, monitoring thinking, and preparing and presenting work with respect to evaluation criteria; (3) culturally influenced processes in solving problems (Hambleton & Murphy, 1992); (4) culturally enriched authentic tasks; (5) low social skills and introverted personality; (6) added communication skills to present, discuss, argue, debate, and verbalize thoughts; (7) inadequate or undue assistance from parents, peers, and teachers; (8) lack of resources inside and outside of schools; (9) incompatibility in language and culture between assessors and students; and 10) subjectivity in rating and informal observations. (A strategy for reducing the influence of extraneous factors in rating that also supports integration of curricula is to employ multiple scales for different attributes embedded in the performance. For example, essays on social studies can be rated on subject matter knowledge, writing quality, and penmanship.)

With equality as the view of fairness, the strategy for reducing bias is to employ judgmental review and statistical analysis to detect and eliminate biased items or tasks. Recognizing the technical difficulties in statistical investigation of bias in performance assessment, Linn et. al. (1991) asserted that "greater reliance on judgmental reviews of performance tasks is inevitable" (p.18).

Equity

Fair assessment that is equitable is tailored to the individual student's instruction context and special background, such as prior knowledge, cultural experience, language proficiency, cognitive style, and interests. Individualization of assessment can be implemented at different levels in the assessment process, ranging from choice of assessment approach (e.g., a project instead of a test), content (e.g., selecting a topic to write an essay on, allowing translation), administration (e.g., flexible time, allowing a dictionary), scoring (e.g., differential weighting), and interpretation (e.g., using a sliding grading scale).

By assessing students using methods and administration procedures most appropriate to them, bias is minimized because construct-irrelevant factors that can inhibit student performance are taken into consideration in the assessment design. For example, in place of a paper-and-pencil word problem test in math to be administered to the class, a teacher could give the test orally to a LEP student, rephrasing the questions and using the student's native language if necessary. When assessment content is customized, congruence between assessment and instruction for all students is enhanced. And, by adjusting scoring and grading procedures individually based on student background and prior achievement, fairness is directly addressed.

Performance assessment, with its ability to provide students with rich, contextualized, and engaging tasks, can allow students to choose or design tasks or questions that are meaningful and interesting to them, can make adjustments based on student experiences and skills, and can test students individually "to insure that the student is fully examined" (Wiggins, 1989, p.708). These characteristics of performance assessment are indeed the major thrusts of equitable assessment. However, it is the individualization strategy and not the performance task that produces bias-free scores. If multiple versions of a multiple-choice test were written for students with varying learning experiences and backgrounds, and the test administered individually with opportunities for students to defend and explain their answers, similar results could be achieved. The persistent gap between majority and minority student performance on accountability tests, even after the introduction of performance-based sections, may be attributable partially to the fact that these tests are standardized.

The major difficulty in individualized performance assessment is assuring comparability of results. Student performance is not consistent across different contexts and topics in writing assessment, and across different experiments and assessment methods in science (see Miller & Legg, 1993). Attempts to develop tasks that are functionally equivalent have been scarce and unsuccessful. For example, it is difficult to construct comparable tasks of equal difficulty in writing assessment (Miller & Legg, 1993); methods of translating a test into another language and establishing the equivalence of scores are not well known and are used sporadically (Hambleton & Kanjee, 1993); and for constructed response exams that allow students to choose a subset of questions, it is not common in tryouts to have representative examinees answering all combinations of the questions (Wainer, Wang, & Thissen, 1994). Procedures for equating scores from disparate assessments are just as problematic. As noted by Linn & Baker (1993), "some desired types of linking for substantially different assessments are simply impossible" (p.2).

Other pitfalls in assuring equity in performance assessment through individualization strategies can also be noted. If students are delegated the responsibility of determining how they should be assessed, such as choosing an essay topic, picking out their best work, or assigning points, individual differences in this metacognitive ability can become a source of bias. Furthermore, for any form of assessment, differential scoring and interpretation (such as the use of differential standards) encourage low expectations for the coddled students, and ultimately lessen their competitive edge when entering the workforce.

Summary

In dealing with the issue of fairness in performance assessment, we are confronted with some dilemmas. On the one hand, assuring equality in performance assessment through standardization enables comparisons of student performance and simplifies administration processes; however, it loses task meaningfulness and creates difficulty in avoiding bias. On the other hand, assuring equity effectively reduces bias and enables rich, meaningful assessment, but it introduces difficulty in administration and in comparing student performance, causes a potential side effect of poorly equipping students for the real world, and can be unfair to students with low awareness of their own abilities and quality of performance. Although standardized assessment is encouraged because it is a requirement for reliability, which is a necessary condition for validity, the hermeneutic approach to score interpretation supports contextualized and non-standardized assessment, and argues that validity can be achieved without reliability (Moss, 1994).

There is currently little research devoted to examining and promoting fairness in performance assessment. However, the urgency to build this knowledge base should not surpass the much needed research on, and efforts to develop, sound and practical performance assessments. When dealing with the issue of fairness in assessment, validity must be considered concurrently. How much better off are we with assessments that are equally invalid for all groups (fair but invalid) than assessments that are invalid for some groups (valid but unfair)?

References

Baker, E. L. & O'Neil, H. F. (1993). Performance assessment and equity. In evaluating education reform: Assessment of student performance. Washington, D.C.: Pelavin Associates.

Hambleton, R. K. & Kanjee, A. (1993). Enhancing the validity of cross-cultural studies: Improvements in instrument translation methods. Paper presented at the annual American Educational Research Association Conference, Atlanta, Georgia.

Hambleton, R. K. & Murphy, E. (1992). A psychometric perspective on authentic measurement. *Applied Measurement in Education, 5*(1), 1-16.

Linn, R. L. (1993). Educational assessment: Expanded expectations and challenge. CSE Technical Report 351, CA: CRESST.

Linn, R. L. & Baker, E. L. (1993). Comparing results from disparate assessments. The CRESST Line, CA.: National Center for Research on Evaluation, Standards, & Student Testing, 1-3.

Linn, R., E., Baker, E. L. & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15-21.

Madaus, G. (1993). A national testing system: Manna from above? An historical/technological perspective. *Educational Assessment, 1*(1), 9-26.

Messick, S. (1989). Validity. In Linn R. (Ed.), Educational Measurement, New York: Macmillan Publishing Company, p. 221-262.

Miller, M. D. & Legg, S. M. (1993). Alternative assessment in a high-stakes environment. Educational Measurement: Issues and Practice, 12(2), 9-15.

Moss, P. A. (1994). Can there be validity without reliability? Educational Researcher, 32(2), 5-12

Wainer, H., Wang, X., & Thissen, D. (1994). How well can we compare scores on test forms that are constructed by examinees' choice? Journal of Educational Measurement, 31(3), 183-199.

Wiggins, (1989, May). A true test: Toward more authentic and equitable assessment. Phi Delta Kappan, 70(9), 703-713.

Tony C. M. Lam is Associate Professor, Faculty of Education, University of Toronto.

ERIC Digest # EDO-CG-95-25

Computer-Assisted Testing in Counseling and Therapy

James P. Sampson, Jr.

Computer-assisted testing (CAT) in counseling and therapy is becoming increasingly common due to dramatic improvements in cost-effectiveness and increased counselor familiarity with computer applications. The assumption underlying the use of CAT is that the effectiveness of counseling is improved by allocating repetitive computational and instructional tasks to the computer, thus allowing counselors to more fully focus on interpersonal tasks, such as helping clients understand the role of testing in counseling and helping clients integrate the self-knowledge obtained in testing into a concrete plan for behavior change. The potential benefits of CAT, however, need to be evaluated against the potential limitations of this technology.

Computer-Assisted Testing Options

The following options exist for using computer-assisted testing in counseling and therapy:

1. Test administration via: (a) keyboard input by the client from items presented on the computer display, with alternative input options available for physically disabled clients; or (b) client completion of a specially prepared test answer sheet that is then optically scanned for computer input; or (c) client completion of a traditional test answer sheet with keyboard input by a clerical staff person.
2. Test scoring via the computer (local or remote).
3. Test score profile generation via the computer.
4. Narrative interpretive report generation via the computer with reports available for both the client and practitioner if appropriate (the narrative report may also include the test profile).
5. Videodisc-based generalized test interpretation provided to the client immediately following test administration (Sampson, 1990a, ph. 452-453).

Potential Benefits of Computer-Assisted Testing

Computer-assisted testing can enhance test administration, scoring, interpretation, and integration. Test administration and scoring may be enhanced due to the standardization inherent in computer functioning. Each test taker receives an identical presentation of test items and response sets (with the exception of adaptive testing where each test taker receives a unique minimum selection of items necessary to achieve a valid result). Greater standardization of item presentation eliminates errors caused when a test taker gets out of sync between the answer sheet and a printed test item (Byers, 1981). The availability of adaptive devices allows persons with a disability to complete tests with minimal staff assistance (Sampson, 1990b). Test results can be more valid since staff members have less of an opportunity to influence client responses. Test scoring is enhanced due to reduced computational errors.

Test interpretation may be enhanced by providing the counselor with an expanded and consistent knowledge base to assist in the interpretation of test data. Computer-based test interpretation (CBTI) is typically based on research data and clinical experience. Roid and Gorsuch (1984) described four approaches to CBTI: (1) descriptive interpretations; (2) clinician-modeled interpretations (renowned clinician type); (3) clinician-modeled interpretations (statistical model type); and (4) clinical actuarial interpretations. Counselors can use CBTI to support or challenge their judgments about the nature of client problems and potentially effective intervention strategies.

Test integration may be enhanced by including computer-assisted instruction as part of CAT. Clients can be better prepared to use their test results by being more aware of basic concepts and the general nature of their scores. Relieved of presenting repetitive test interpretation information, counselors have more time to explore clients' perceptions of their test data and the implications of the test data for behavior change. The computer can be used to deliver both text-based and interactive video-based instruction (Sampson, 1990a).

Potential Limitations of Computer-Assisted Testing

Computer-assisted testing can limit, as well as enhance, test administration and interpretation. Although paper-and-pencil and computer administration of tests often produce equivalent results, variations in results have sometimes been found to exist. French (1986) recommended that the equivalency of results from different types of administration modes needs to be established for each instrument. Establishing equivalency will reduce the likelihood that computer administration is influencing the nature of test results. Scoring errors represent another potential limitation for computer-assisted test administration. Most (1987) noted that, "The computer itself does not contribute error, but the complex nature of computer programming and the difficulty involved in reading computer programs or code makes it easy to make program errors which are difficult to find" (p. 377).

Concerns have been raised about the validity of computer-based test interpretation. Eyde and Kowal (1987) found differences in CBTI reports generated from a single set of scores from one instrument. Differences also were noted in their study between the CBTI reports and the judgments of a clinician. Eyde and Kowal (1987) stated, "Buyers should be aware of the limitations of computer products and remind themselves that computer output is only as good as the data behind the decision rules used to produce the interpretation" (p. 407). Ethical concerns also exist about counselor misuse of CBTI. Unqualified counselors may be more likely to use CBTI reports to compensate for a lack of training and experience. By using CBTI to replace rather than supplement counselor judgment, counselors become more dependent on the potentially dubious validity of some CBTI software and are less likely to integrate data from valid CBTI reports effectively with other sources of client data due to their lack of background knowledge.

Recommendations

Counselors should become familiar with existing CAT applications (see Krug, 1993; Walz, Bleuer, & Maze, 1989) and the various professionalstandards that relate to CAT. Counselors then should carefully select and effectively implement valid software that is subsequently evaluated in terms of service delivery impact.

Conclusion

The use of CAT can either enhance or limit the effectiveness of testing in counseling and therapy. Having an open mind about the potential of this technology and a willingness to change needs to be matched with good critical thinking skills and a healthy skepticism for any innovation promising substantial benefits from minimal efforts. It is the responsibility of counselors to guide the design and use of this technology.

References

Byers, A. P. (1981). Psychological evaluation by means of an on-line computer. Behavior Research Methods & Instrumentation, 13, 585-587.

Eyde, L. D., & Kowal, D. M. (1987). Computerised test interpretation services: Ethical and professional concerns regarding U.S. producers and users. Applied Psychology: An International Review, 36(3/4), 401-417.

French, C. F. (1986). Microcomputers and psychometric assessment. British Journal of Guidance and counselling, 14, 33-45.

Krug, S. E. (1993). PSYCHWARE SOURCEBOOK: A resource guide to computer based products for assessment in psychology, education, and business (4th Ed.). Champaign, IL: Metritech.

Most, R. (1987). Levels of error in computerized psychological inventories. Applied Psychology: An International Review, 36(3/4), 375-383.

Roid, G. H., & Gorsuch, R. L. (1984). Development and clinical use of test-interpretive programs on microcomputers. In M. D. Schwartz (Ed.), Using computers in clinical practice: Psychotherapy and mental health applications (pp. 141-149). New York: Haworth Press.

Sampson, J. P., Jr. (1990a). Computer applications and issues in using tests in counseling. In C. E. Watkins, Jr. & V. L. Campbell (Eds.), Testing in counseling practice (pp. 451-474). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Sampson, J. P., Jr. (1990b). Computer-assisted testing and the goals of counseling psychology. The Counseling Psychologist, 18, 227-239.

Walz, G. R., Bleuer, J. C., & Maze, M. (Eds). (1989). Counseling software guide: A resource for the guidance and human development professions. Alexandria, VA: American Association for Counseling and Development.

James P. Sampson, Jr. is a professor in the Department of Human Resources and Studies and Co-Director of the Center for the Study of Technology in Counseling and Career Development at Florida State University in Tallahassee, Florida.

ERIC Digest # EDO-CG-95-26

Testing Students With Disabilities

Kurt F. Geisinger and Janet F. Carlson

The assessment of students with disabilities has taken on considerable importance since the passing of the Americans with Disabilities Act (ADA) of 1990, although most of the requirements for assessing students were previously justified legally based upon Section 504 of the Rehabilitation Act of 1973. Generally, the best methods for assessing students with disabilities coincide with legally defensible methods for this activity. Under ADA, a "disability is defined as (a) a physical or mental impairment that substantially limits one or more life activities, (b) a record of such an impairment, or (c) being regarded has having an impairment despite whether or not the impairment substantially limits major life activities" (Geisinger, 1994, p. 123). ADA requires that assessment of individuals with disabilities be performed with any reasonable accommodations being made. The word, "reasonable," of course, is ambiguous and differs depending upon the circumstances of the assessment.

The considerations involved in assessing students with disabilities are presented below under three related activities: test selection, test administration, and test interpretation. Additional considerations are noted at the conclusion of this digest

Test Selection

When counselors assess either an individual with a disability or a group of individuals including those with disabilities, we must consider test selection. We must ask questions regarding prospective instruments that address the assessment's suitability for use with students with disabilities. Most important, we should consider whether individuals with like disabilities were included in the normative and validation samples. Whether there are specialized administrative procedures and forms, such as large-type test forms for individuals with visual disabilities or untimed administrations for individuals with learning disabilities, is also important.

In some cases, where no measures of a given attribute are available with adapted administrations, a counselor might consider how easily he or she can adjust an instrument for use with test takers who have a disability. When published instruments are adapted, however, interpretations of the results should be tentative. Should there be planned test administrations or specialized administrative procedures for those with common disabilities, we also need to determine whether the testing instrument has norms available for those with various common disabilities—or in specialized cases, for the specific disability with which the counselor is concerned. Are there parallel interpretive guides for evaluating the assessment results for those with specific disabilities and for those who have taken specialized administrations of the assessment? Finally, are these specialized interpretive guides, if available, based upon empirical reliability and validation research?

If positive answers to the above questions are not found, a counselor should consider whether the use of an unvalidated instrument is justified. When counselors adapt a measure themselves (e.g., reading an assessment to a test taker when normal administration calls for the test taker to read the test questions), they are essentially using an unvalidated instrument. As such, we must ask whether this instrument is likely to yield usefulinformation over and above that which is already available from non-test sources. The answer to this question is likely to differ based upon the nature of the decision for which the assessment is being used.

Test Administration

The most important question for a counselor regarding test administration to a student with a disability is whether the student can be appropriately and meaningfully assessed using the conditions under which the instrument was standardized. We should consider students' backgrounds, skills, abilities, and other characteristics if we are unsure. If such an evaluation does not answer the question adequately, one should seek advice from colleagues or the test publisher. Professionals who work frequently with students with disabilities, such as special educators, may be especially helpful, even if they are not experts on assessment. Ask such individuals about the kinds of tasks students with the kinds of disabilities, backgrounds, skills, abilities, and other characteristics are able to perform. Then evaluate the test materials against these tasks. It may be especially helpful to talk with professionals who know the student. When talking with professionals who do not know the individual, provide an assessment of the degree of severity of the disability.

Some assessments offer specialized administrations for individuals with common types of disabilities. These assessments tend to be either those oriented specifically for use with students with disabilities or those in widespread use, such as frequently used college admissions measures. Some accommodations permit continued administration in group settings; others require individual administration. For example, assessments may be available in improved-type, large-type, Braille, and audiocassette versions for those with visual disabilities. "Time limits can be enforced, extended, or waived altogether. Test takers may be given extra rest pauses, a reader, an amanuensis (a recorder), a sign language interpreter, a tape recorder to register answers, convenient test taking locations and assessment times, and other accommodations as needed to meet their particular requirements" (Geisinger, 1994, p. 124). Accessibility to the assessment site also needs to be considered.

Under rare circumstances, it may be necessary for a counselor to adapt a professionally developed assessment device for administration to a specific student. Such procedures should be performed only when no valid measure exists for the given assessment. If a counselor makes an adaptation, he or she must be aware that the scoring, norms, and interpretation are compromised and cannot be used validly. To the extent that the adaptation is extremely minor, of course, it may fall within normal variation of test administrations. However, any serious adaptation does jeopardize the value of using a published measure.

Test Interpretation

When interpreting the results of an assessment of a student with a disability who nevertheless took the assessment under standardized conditions, we can employ the normal judgment process, although we also should follow any advice provided in the test manual. It is particularly advisable to check whether any validation studies using populations including students with the disability in question have been performed. Similar caveats apply when employing a standardized adaptation, such as an untimed administration or the use of a Braille version.

When a counselor has performed an adaptation of an assessment or uses a locally derived adaptation, then extreme caution should rule as far as test interpretation is concerned. The modified assessment simply is not the same measure as the original version for which norms and validation results exist. In general, results from such a measure are best interpreted by developing hypotheses as opposed to making decisions (Phillips, 1994).

The goal of any interpretation of a modified assessment should be an expected result on the comparable standardized assessment. "We wish to know how the person taking an adapted form of a test would have performed if he or she could have taken the test under standardized conditions, assuming that the disabilities did not exist" (Tenopyr, Angoff, Butcher, Geisinger, & Reilly, 1993, p. 2).

Additional Issues

Several special issues related to the assessment of students with disabilities deserve mention. First, some information on the extent and severity of a student's disability should be acquired before an assessment either is selected or administered. Such information may help guide the counselor in making these decisions.

It also may be appropriate to choose and administer measures that assess compensatory skills used by persons with disabilities. It makes little sense, for example, to administer an assessment of graph-reading ability to a student with a severe visual disability. It would be more useful to determine how such students consider graphical information (e.g., via textual analysis with material written in Braille) and provide a direct assessment thereof.

Those purchasing assessment instruments should carefully evaluate all measures to determine the degree to which they have been used with and adapted for students with disabilities. If one is disappointed with the robustness of a measure (Geisinger, 1994) when it is used with students with disabilities, let the publisher know. With enough input, they may become more interested in making needed changes. Relatedly, when one discovers measures, administrative modifications, or interpretive strategies that are well-suited for use with students with disabilities, share the results. Such findings are too important to keep secret.

References

Americans with Disabilities Act of 1990, 42 U.S.C. § 12101 et seq (1990).

Geisinger, K. F. (1994). Psychometric issues in testing students with disabilities. Applied Measurement in Education, 7, 121-140.

Phillips, S. E. (1994). High-stakes testing accommodations: Validity versus disabled rights. Applied Measurement in Education, 7, 93-120.

Section 504 of the Rehabilitation Act of 1973, 29 U.S.C. § 701 et seq (1973).

Tenopyr, M. L., Angoff, W. H., Butcher, J. N., Geisinger, K. F., & Reilly, R. R. (1993). Psychometric and assessment issues raised by the Americans with Disabilities Act (ADA). The Score, 15(4), 1-2, 7-15.

Kurt F. Geisinger is Dean of the College of Arts and Sciences and Professor of Psychology at the State University of New York at Oswego.

Janet F. Carlson is Assistant Professor of Counseling and Psychological Services at the State University of New York at Oswego.

ERIC Digest # EDO-CG-95-27

The School Psychologist's Role in School Assessment

Sylvia Rosenfield and Deborah Nelson

Psychological services for children originated within a diagnostic testing model. Psychometric techniques were developed to assess individual children's cognitive-intellectual, personality, and academic functioning. Today, testing techniques have achieved a high degree of prominence and testing is a major industry.

Recently, however, assessment in the field of school psychology has been changing and reshaping itself to meet the demands of public policy and litigation, the requirements of an increasingly diverse student population, and the constant shifting of educational concerns. There have been, as well, continual refinements in the concepts and technology of measurement (Taylor, Tindal, Fuchs, & Bryant, 1993). These changes have challenged all school professionals to modify their assessment practices in order to adapt to them. However, within the schools, it remains true that there are few others with training, experience and expertise in assessment comparable to that of school psychologists.

Traditionally, school psychology has emphasized diagnosis and classification of individual students, and school psychologists have acted as gatekeepers for special services. But as the current ethical, political, legal, and educational context has evolved, there has been a re-examination of the purposes and applications of data gathered during the assessment process (Taylor, et al., 1993). In a position paper, *The Role of the School Psychologist in Assessment* (1994), the National Association of School Psychologists endorsed the proposition that assessment practices must be linked to prevention and intervention to provide positive outcomes for students. Thus, there is an increasing emphasis on information that is "useful in designing, implementing, monitoring, and evaluating interventions" (Reschly, Kicklighter, & McKee, 1988). Moreover, it is suggested that school psychologists assist both local education agencies and state education agencies in restructuring schools in positive ways. One of the constant elements in the school restructuring movement is the call for greater accountability at every level, which has resulted in "innovative thinking about alternative forms of assessment" (Stiggins & Conklin, 1992, p. 3).

This broader, more outcome-based approach to the use of assessment in schools has had an impact on the assessment practices of school psychologists. Currently, there are at least three major purposes of school psychological assessment: informing entitlement/classification decisions, planning interventions, and evaluating outcomes.

Assessment Purposes

Entitlement/Classification Decisions

Although, historically, the school psychologist has been the professional to develop an individual diagnosis of a referred student using psychoeducational tests, that role became even more routinized as a result of the 1975 federal legislation, P.L. 94-142, requiring testing for classification prior to delivering services to children with handicapping conditions. However, there have been recent changes in the field of special education, with pressure increasing for inclusive placements in regular education classrooms even for students with severe and profound disabilities. These pressures arose from research demonstrating limitations of the traditional classification, labeling, and placement procedures, many of which relied upon school psychologists' testing of students referred for problems. Challenges to the norm-referenced tests used to justify the classification and placement decisions arose for many reasons, including "lack of data to support the use of certain types of tests..., litigation related to the discriminatory nature of other types..., and the general feeling that most tests did not provide educationally relevant information" (Taylor, et al., 1993, p. 114).

Since federal law and related state regulations still, in most cases, require labeling for funding purposes, norm-referenced psychoeducational assessment will likely continue in the schools to fulfill the legal mandate. However, currently there is an emphasis upon improving the technical characteristics of the most commonly used tests to answer growing concerns about the soundness of many of these instruments. In addition, several basic constructs underlying these tests have been revised, and new constructs of cognition and neuropsychological and psychological processes, such as memory and metacognition, are finding their way into new test construction and revisions of older instruments (Taylor, et al., 1993). How useful these new and revised tests and their underlying constructs are remains open for further study, although there continue to be weak or nonexistent links to interventions for most psychoeducational tests (Macmann & Barnett, 1994). In addition, as requirements for eligibility for funding are modified, the use of tests for these purposes will also evolve

Assessment Linked to Intervention

Perhaps the most far-reaching change in the role for school psychologists has been an increased emphasis on linking assessment and intervention, so that information from the assessment process leads directly to intervention strategies rather than just to a diagnostic label and alternative placement for the student. School psychologists have moved from relying upon standardized/norm-referenced testing practices to frequent use of more natural and dynamic forms of assessment that impact directly on classroom instructional delivery and behavior management. The importance of this shift arises from the current state of classroom assessment. While the instructional and management decisions that teachers make about their students have been recognized as critical to important outcomes, relatively little attention has been paid to the quality and process of classroom assessment in research or practice. This has been true in spite of evidence that teachers are concerned about the quality of their own assessments, and have limited knowledge of assessment methodologies and their use in instructional decision making (Stiggins & Conklin, 1992). Increasingly, school psychologists have become involved in developing and delivering behavioral and curriculum-based assessment procedures useful for classroom decision making to assist teachers.

A recent development has been the growth of curriculum-based assessment methods that use direct observation and recording of student performance in the classroom curriculum itself to gather information for instructional decisions. Two major forms of this type of assessment are the curriculum-based assessment for instructional design (CBA-ID) model (e.g., Gickling & Rosenfield, in press), and the curriculum-based measurement (CBM) model (e.g., Deno, 1986). CBA-ID was designed to assist teachers in planning instruction for individual students, whereas CBM was developed primarily to assess pupil progress in the classroom. The information derived from these techniques is used by school psychologists consulting with teachers to support them in developing interventions related to instruction and classroom management (Rosenfield, 1987). These classroom-based models of assessment are also used by prereferral and support teams designed to provide assistance to teachers and students

Outcome Evaluation

School reform has created a focus on the outcomes of education. Psychologists are involved in discussions of a possible national test to be given to all students, and state assessments aligned with state content standards are in the process of development. Many of these will be performance assessments, which still have serious technical issues that need to be resolved (Ysseldyke, 1994). School psychologists have a role in helping school personnel understand and use the results of these external assessments.

At the local level, outcome assessment is also changing. Reform in regular and special education often involves the creation of new programs. School psychologists can bring their assessment expertise to the school reform agenda by helping school systems and individual schools evaluate the effectiveness of different programs and organizational changes designed to meet specific goals. School psychologists can provide assistance in systems change efforts, including needs assessment prior to program implementation, as well as on-going monitoring of program implementation and effectiveness along a broad array of outcome dimensions, depending upon the goals of the school personnel. Conducting research and evaluation to answer important questions about effective programs is an additional assessment role in which many school psychologists can participate.

Summary

School psychologists can play a unique role in schools because of their assessment expertise. Traditionally, they have been most involved in individual psychoeducational assessment for classification and labeling purposes, but the limitations of this form of assessment for building intervention strategies have led many school psychologists to broaden their role. Techniques linking assessment to interventions are being demonstrated by school psychologists as they consult with teachers to enhance the classroom performance of students. Further, school reform initiatives have required more program evaluation at the building and system levels, and school psychologists are engaged in these activities as well. Assessment is an important task in the schools, and school psychologists can increase their impact on school effectiveness by contributing their expertise in this domain at many levels.

References

Deno, S.L. (1986). Formative evaluation of individual programs: A new role for school psychologists.*School Psychology Review, 15*, 358-374.

Gickling, E., & Rosenfield, S. (in press). Best practices in curriculum-based assessment. In A. Thomas & J. Grimes (Eds.). *Best practices in school psychology,* Vol. 3. Washington, D.C.: National Association of School Psychologists.

Macmann, G.M., & Barnett, D.W. (1994). Structural analysis of correlated factors: Lessons from the verbal-performance dichotomy of the Wechsler Scales. *School Psychology Quarterly, 9,* 161-197.

National Association of School Psychologists (1994). *The role of the school psychologist in assessment.* Washington, DC: Author.

Reschly, D.J., Kicklighter, R., & McKee, W. (1988). *School Psychology Review, 17*(1), 9-50.

Rosenfield, S. (1987). *Instructional consultation.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Stiggins, R.J., & Conklin, N.F. (1992). *In teachers' hands.* Albany, NY: State University of New York Press.

Taylor, R.L., Tindal, G., Fuchs, L., & Bryant, B.R. (1993). Assessment in the nineties: A possible glance into the future. *Diagnostique, 18,* 113-122.

Ysseldyke, J. (1994). Assessment: Current directions and misdirections. Unpublished paper.

Sylvia Rosenfield is professor and chair of the Department of Counseling and Personnel Services at the University of Maryland--College Park.

Deborah Nelson is a doctoral student in the school psychology program at the University of Maryland--College Park.

ERIC Digest # EDO-CG-95-28

Cooperation Between School Psychologists and Counselors in Assessment

Douglas K. Smith

The role of school psychologists and counselors in assessment is well established and is a frequent research topic. For example, a review of the ERIC database from 1987 to 1994 revealed 64 entries for "assessment and school psychology" and 622 entries for "assessment and counseling." Similar results were obtained for a review of the Psychological Abstracts database with 146 entries for "assessment and school psychology" and 924 entries for "assessment and counseling." However, studies that explored the joint role of counselors and school psychologists in assessment could not be located. With the current emphasis on collaboration in schools and the use of a pupil services model to deliver services of counselors, school psychologists, school social workers, and school nurses, it is important to examine ways in which school psychologists and counselors can work together in the assessment process.

School Psychologists and Assessment

While the assessment activities of school psychologists emphasize services to children and youth, usually within a school setting, the assessment activities of counselors frequently cover a wider age range and emphasize the adult population. The assessment of individual students is both the traditional and the major role of school psychologists (Fagan & Wise, 1994). In fact, surveys of school psychologists continue to show that the majority of their time is centered on assessment activities. A recent survey (Smith, Clifford, Hesley, & Leifgren, 1992) indicated that the typical school psychologist devoted 53% of his or her time to assessment with the assessment of intellectual ability being the primary focus. Techniques that are used emphasize structured, standardized formats with an emphasis on quantitative rather than qualitative approaches (Smith & Mealy, 1988; Smith, Clifford, Hesley, & Leifgren, 1992). In general, the school psychologist's involvement in assessment begins with a student who has been referred by a parent or teacher for academic or behavioral difficulties. As part of the assessment process, an individual test of intelligence and an achievement test are likely to be administered. Additional data that are collected may include behavioral observations, rating scales completed by teachers and parents, and interviews with the student and with the student's parents and teachers.

Counselors and Assessment

As Hood and Johnson (1991) note, "Assessment is an integral part of counseling...[and] provides information that can be used in each step of the problem-solving model" (p. 11). In general, assessment information is used to clarify concerns of clients, to plan programs or interventions and evaluate their effectiveness, to provide career planning information, and to assist clients in undertanding themselves. Thus, counselors, especially in school settings, are more likely than school psychologists to be involved in developmental assessment approaches that are holistic in nature, are qualitative rather than quantitative, and emphasize developmental norms. These approaches may include checklists or rating scales, unfinished sentences, writing activities, decision-making dilemmas, games, art activities, story-telling and bibliotherapy techniques, self-monitoring techniques, role-play activities and play therapy strategies (Vernon, 1993). Surveys of counselors in different counseling settings including counseling agencies, secondary schools, and private practice indicate that counselors use a variety of test instruments with an emphasis on interest inventories, personality inventories, and aptitude tests (Hood & Johnson, 1991).

Both school psychologists and counselors are involved in the assessment process with differing emphases and orientations that are complementary to each other. School psychologists often emphasize the use of quantitative approaches to measure ability and academic skills while counselors often utilize developmental as well as qualitative approaches to assess personality characteristics, interests, and aptitudes. The two approaches, when combined, can offer a more comprehensive picture of a student than either approach alone.

Multidisciplinary Teams and Collaboration

With the advent of Public Law 94-142 (the Education of All Handicapped Children Act) and the Individuals with Disabilities Education Act (IDEA), emphasis was placed on a multidisciplinary approach to assessment and placement activities for students referred for possible disabilities. Multiple sources of information, multiple procedures, and multiple settings are required in order to develop a comprehensive understanding of students' needs and abilities. The basis for such an approach is collaboration among professionals including regular education teachers, special education teachers, administrators, pupil services personnel, and parents.

Collaboration, of course, is not a new concept. Sullivan (1993) describes it as "a reform movement that has been gaining in momentum over the past five years" (p. 1) and suggests that it was created as a response to the fragmentation in service delivery that often occurs in educational and mental health settings. Both benefits and obstacles are associated with collaboration. A major benefit of collaboration is the opportunity to create a more comprehensive approach to service delivery. It facilitates development and sharing of new perspectives on how students can be served and promotes improved communication among those working with students. Collaboration can also foster an emphasis on prevention and can create more effective services by reducing duplication. In order for collaboration to be successful, however, it must receive support at all levels and participants must display cooperation and trust (Sullivan, 1993).

Recommendations for Collaboration

Counselors and school psychologists have much to offer in the assessment of students and both sets of professionals should be members of multidisciplinary assessment teams. Counselors contribute skill in developmental assessment approaches and provide a holistic view of the student. In addition, their expertise in interpersonal assessment and career/vocational assessment is valuable in program planning, especially for adolescents. School psychologists' contributions include expertise in the assessment of cognitive and academic skills and the development of classroom interventions. Their background in behavior management and educational psychology along with training in psychological assessment provide a unique perspective for program planning.

The increased focus on involving families in prevention and intervention programs offers counselors and school psychologists the opportunity to collaborate in a number of ways. Activities in which the two sets of professionals can work together include family counseling, parent training, and the development and implementation of behavior management programs in the home. The assessment skills of both specialties can also be utilized to develop evaluation procedures to examine the effectiveness of programs.

Within the school setting itself, a number of opportunities exist for counselors and school psychologists to work together. These include developing support groups for students, working with classroom teachers to implement developmental guidance materials and curriculum within the classroom, and developing aggression/violence prevention programs and curricula. By utilizing the unique assessment training and expertise of counselors and school psychologists we can develop a more accurate picture of the whole student and his or her specific needs. In this way more effective intervention and prevention programs can be developed and implemented.

Summary

Both counselors and school psychologists are trained in assessment with somewhat differing emphases and areas of expertise. The multidisciplinary approach to assessment required by P. L. 94-142 and IDEA is especially suited for the two groups of assessment professionals to work together in a collaborative manner. In this way a more complete picture of students' needs can be developed and service delivery can be enhanced.

References

Fagan, T. K., & Wise, P. S. (1994). *School psychology: Past, present, and future*. New York: Longman.

Hood, A., & Johnson, R. (1991). *Assessment in counseling: A guide to the use of psychological assessment procedures*. Alexandria, VA: American Association for Counseling and Development.

Smith, D. K., Clifford, E. S., Hesley, J., & Leifgren, M. (1992). *The school psychologist of 1991: A survey of practitioners*. Paper presented at the Annual Meeting of the National Association of School Psychologists, Nashville, TN.

Smith, D. K., & Mealy, N. S. (1988). *Changes in school psychology practice: A five year update*. Paper presented at the Annual Meeting of American Psychological Association, Atlanta. (ERIC Document Reproduction Service No. ED 302 782).

Sullivan, D. (1993). *Benefits and obstacles to collaboration*. Madison, WI: Bureau for Pupil Services, Wisconsin Department of Public Instruction.

Vernon, A. (1993). *Developmental assessment and intervention with children and adolescents*. Alexandria, VA: American Counseling Association.

Douglas K. Smith, Ph.D., NCSP, is Professor of School Psychology, Chairperson of the Department of Counseling and School Psychology, and director of the School Psychology Program at the University of Wisconsin-River Falls.

ERIC Digest # DO-CG-95-29

Using Buros Institute of Mental Measurements Materials in Counseling and Therapy

Barbara S. Plake and Jane Close Conoley

Assessment use is a cornerstone of successful counseling. Information from assessments is used for making initial diagnostic decisions, to assess client readiness for clinical interventions, for monitoring progress during the counseling process, and for assessing therapeutic outcomes at the conclusion of the counseling program. Therefore, counselors' needs for information about assessment devices and approaches are very high. Tests are being published at a remarkable rate; it is a challenge for the practicing counselor to stay well informed about new approaches and revisions of well-known tests (APA, 1990; Rudner and Dorko, 1989).

Buros Institute of Mental Measurements

Established over 50 years ago by Oscar K. Buros, the mission of the Buros Institute is to improve tests and testing practices by providing candidly critical reviews of instruments. The Institute fulfills this mission, in part, by publishing several reference works that contain descriptive and quality information about commercially available tests. In addition, the Institute publishes a topical series and sponsors symposia on specific assessment areas. Through access to products and programs of the Buros Institute, counselors can make more informed assessment selection decisions and stay current with assessment practices in the field.

Mental Measurements Yearbook and Tests in Print Series

The *Mental Measurements Yearbook* (MMY) and *Tests in Print* (TIP) series serve as companion resources for locating and evaluating commercially available tests. The MMY is hierarchically organized. Each new volume contains information about new or revised tests made available since the last publication. TIP, on the other hand, is comprehensive, providing descriptive information about currently available tests. In addition, TIP is a cumulative index to test information in the MMY series.

The Mental Measurements Yearbook contains descriptive information about tests, including test author, publisher, publication date (including dates of revisions), purpose of the test, categorization of the test, target populations/age ranges, price lists, reported scores, and availability information for users' documents such as manuals. For most tests, evaluations of the quality and utility follow descriptive information. Typically, two independent reviewers prepare critical analyses. Another useful feature of the information in the *Mental Measurements Yearbook* is the accompanying listing of journal references associated with each test.

Uses of the MMY Series for Counseling Practices

Information contained in the MMY can aid counselors in many ways (Claiborn,1991; Plake, Conoley, Kramer & Murphy, 1991). By referencing the categorization system developed by the Institute, counselors can locate tests appropriate for their purpose. These categories indicate the test's general purpose, such as Achievement, Behavior Assessment, Developmental, Education, Fine Arts, Intelligence and Scholastic Aptitude, Mathematics, Multi-Aptitude Batteries, Neuropsychological, Personality, Reading, Science, Sensory-Motor, Social Studies, Speech and Hearing, and Vocations.

Tests can also be identified through the score index. If, for example, a counselor wants to assess client self-esteem, referencing the score index will yield tests that provide a self-esteem score. The descriptive and evaluative information for those particular tests in the Yearbook will assist the counselor to identify the possible assessment instruments best suited for the client characteristics and clinical purpose.

The critical evaluations prepared by independent expert reviewers also assist the counselor by providing thorough, thoughtful analyses of the quality and utility of the test. Not only are these reviews helpful in making informed decisions about the usefulness of the test for the particular situation, but they also communicate current thinking in the field about the construct the test is designed to assess. Therefore, the reviews serve a continuing education purpose for practicing counselors by assisting the counselor in keeping current with theory and assessment developments.

It is sometimes important for counselors to be able to articulate and defend their assessment choices to a variety of audiences. The descriptive and evaluative information about an instrument, prepared by experts in the field, can serve as a definitive reference that has been shown to be useful for legal purposes. Although not designed for litigation, these reviews are potentially useful in court hearings when questions pertaining to assessment selection are raised. In addition, these reviews can provide objective evidence for such purposes as quality assurance reports and evaluations or audits of counseling practices or programs.

Other Products and Programs from the Buros Institute

In addition to the *Mental Measurements Yearbook* and *Tests in Print Series,* the Institute sponsors other products and programs that are potentially valuable for counselors and therapists.

Buros-Nebraska Symposium and Series on Measurement and Testing. One notable program is the Buros-Nebraska Symposium on Measurement and Testing. At these symposia, key people in the field are invited to make presentations and to lead discussions on issues relevant to assessment. The Buros-Nebraska Symposium Series is approved by American Psychological Association as a sponsor of continuing education credits. Counselors can acquire APA-approved CEUs through attendance. The presentations are edited and produced into volumes that are published in the Buros-Nebraska Series on Measurement and Testing. Occasionally, additional chapters are included in the volumes in order to more fully represent the topical area. Two recent symposia are of particular relevance to counseling practice: Family Assessment and Multicultural Assessment.

Oscar K. Buros Library of Mental Measurements. Located at the University of Nebraska, the Oscar K. Buros Library of Mental Measurements is also a useful resource for counselors and therapists. Counselors can inspect tests and ascertain their appropriateness for particular clinical purposes before purchase. The tests reviewed for the MMY and TIP series are located in the Institute's library and are available for public inspection. Tight restrictions are placed on access for secure tests and all copyright materials are protected from dissemination through strict policies and procedures. However, the library is a significant resource of tests useful for counseling purposes.

Electronic Access to Test Review Information. The Eleventh MMY was also produced as a CD-ROM, searchable both through the traditional indices and by search algorithms. This product also provides a comprehensive master index to the location of test information in the Mental Measurements Yearbook and Tests in Print Series. The Institute is investigating other options for providing electronic access to test information.

Buros Desk Reference (BDR) Series. A new product from the Institute, the BDR is targeted for the individual practitioner. Descriptive and evaluative information about tests most frequently used in particular fields is located in a single volume. The first product in this series, Buros Desk Reference :Psychological Assessment in the Schools, contains evaluative information for over 100 tests most frequently used by school psychologists, counselors, and counseling psychologists.

Summary

Products and programs from the Buros Institute of Mental Measurements serve test information needs of counselors and therapists. The *Mental Measurements Yearbook* and *Tests in Print series* contain information about availability, quality, and utility of assessment devices. Counselors can identify tests potentially appropriate for their clinical practice and stay up-to-date on assessment of psychological constructs and educational outcomes through use of these volumes. In addition, the Institute sponsors topical symposia and volumes targeted at specific audiences; these can provide cutting-edge assessment information to counselors and therapists. Test users can also inspect instruments on-site at the Buros Library of Mental Measurements.

The counseling process is multifaceted and complex. Tests and other specific assessment approaches are useful in assisting counselors in making appropriate clinical decisions; the Buros Institute's mission is to support well-informed assessment decisions. The Institute's products and programs point counselors toward reliable, valid, state-of-the art measurement practice in efficient, effective ways. In this way the counselor's goal to serve the client is enhanced.

Further information about the Institute or any of the products mentioned in this digest is available by writing to the Buros Institute, 135 Bancroft Hall, University of Nebraska, Lincoln, NE 68588-0348.

References

American Psychological Association, (1990). *Finding information about psychological tests: A guide for locating and using both published and unpublished tests.* Washington, DC: Author, Science Directorate.

Claiborn, C. D. (1991). *The buros tradition and the counseling profession.* Journal of Counseling and Development, 69, 456-457.

Plake, B. S., Conoley, J. C., Kramer, J. J., & Murphy, L. U. (1991). The Buros Institute of Mental Measurements: Commitment to the tradition of excellence. *Journal of Counseling and Development*, 69, 449-455.

Rudner, L. M. & Dorko, K. (1989). *Finding information about standardized tests.* (Eric Clearinghouse on Tests, Measurement, and Evaluation Digest 2.)

Barbara S. Plake, Ph.D., is the director of the Buros Institute of Mental Measurement and the W.C. Meierhenry Distinguished Professor of Educational Psychology at the University of Nebraska-Lincoln.

Jane Close Conoley, Ph.D. is an Editor of the Mental Measurements Yearbook series and is a Professor of Educational Psychology and Associate Dean for Research at the Teachers College, University of Nebraska-Lincoln.

ERIC Digest # EDO-CG-95-30

Internet Resources for Guidance Personnel

Liselle Drake and Lawrence M. Rudner

With tens of thousands of information providers and millions of users, the Internet is an enormous and growing resource for guidance counselors and other personnel service professionals. The task for busy professionals is to be able to rapidly identify Internet resources so they can be efficiently incorporated in their work.

In this digest, we identify Internet resources of particular interest to the guidance community. In particular, we identify list servs of interest, identify the offerings of several gopher sites, show you how to access ERIC on the Internet, and describe the AskERIC e-mail service. The novice user is referred to the excellent, well-known electronic books referenced at the end of this digest.

Listservs

Listservs are electronically facilitated discussion forums of participants who share a common interest. You e-mail a thought, question, or response to the forum and the listserv software re-transmits your e-mail to the entire mailing list.

Some listservs of interest to guidance personnel are:

ICN - International Counselor Network - Comprised of counselors, counselor educators, teachers, and graduate students, this listserv contains discussions regardingmental health issues. Subscribe to: ICN-request@ctrvax.vanderbilt.edu

YOUTHNET - This is a discussion list for therapists and other service providers working with youth. Subscribe to: listserv@indycms.iupui.edu

BEHAVIOR - Behavioral and Emotional Disorders in Children - This list discusses psychological disorders in children. Subscribe to: listserv@asuacad.bitnet

VOC-NET-This is the U.C. Berkeley general discussion list and bulletin board on current trends in vocational education. Subscribe to: listserv@cmsa.berkeley.edu

DSSHE-L - Disabled Student Services in Higher Education - The scope of this list encompasses career counseling for students with disabilities; the removal of barriers, both architectural and attitudinal, for the disabled; testing and other academic accommodations; and legal issues pertaining to the Americans with Disabilities Act. It is appropriate for all counselors. Subscribe to: listserv@ubvm.cc.buffalo.edu

TRDEV-L - Training and Development Discussion List - This forum for the exchange of information on the training and development of human resources aims to stimulate research collaboration and assistance in T&D for the professional and academic communities. Subscribe to: listserv@psuvm.psu.edu

AERA-E - American Educational Research Association Division E: Counseling and Human Development- This forum discusses recent research and research ideas. Subscribe to: listserv@asuacad.bitnet

To subscribe to a listserv, send e-mail to the listserv address with a one-line message: SUB listname your-name. For example, to subscribe to AERA-E, you would send e-mail to listserv@asuacad.bitnet. The one-line message would be SUB AERA-E Liselle Drake. To unsubscribe, the message would be UNSUB AERA-E.

Gopher Sites

Gophers are menu-driven systems providing access to a wide range of information. Via Internet and Gopher software, you literally connect to computers across the world to obtain information. Often the information is in the form of text files. The information can also be in the form of searchable databases, directories, software, and graphics. Some Gopher sites of interest to the guidance community include:

Arizona State University - Containing the largest listing of educationally relevant gophers, the ASU Gopher is an excellent starting point. There are pointers to the Best of the Internet for Educators and to a large assortment of electronic journals and newsletters, including Journal of Counseling and Development, Journal of Distance Education and Communication, Psychology, and Rasch Measurement Transactions. Also, by following the path to Electronic journals at CICNET/Electronic Serials/, the counselor can access the full-text issues of the journal, Conflict Resolution Consortium. Gopher to info.asu.edu, select ASU Campus Wide Information/ College of Education/Electronic Journals/.

National Parent Information Network - NPIN provides information and communication capabilities to parents and those who work with them. NPIN offers full-text documents, brochures and other publications which it has gathered from all ERIC components, the National Urban League, the Illinois Parent Initiative, the National PTA, the North Central Regional Education Laboratory, and the Family Literacy Center. NPIN's "Short Items for Parents" are equally suitable for parent educators and counselors in that their scope encompasses issues of discipline, self-esteem, special needs and health of children according to the children's age and developmental levels. The weekly newsletter "Parent News" is another notable offering. Gopher to gopher.prairienet.org, select Education/ERIC/NPIN/.

The U.S. Department of Education's Gopher - USDE posts an extremely large collection of full-text material and information. Of particular interest is A Teacher's Guide to the U.S. Dept. of Education, with selections for "Services and Resources" which offers resources for Drug FreeSchools and Communities, Bilingual Education/Minority Language Affairs, Special Education Programs and Literacy Resource Centers. Gopher to gopher.ed.gov.

The Child, Youth and Family Education and Resources Network - CYFERNET describes its mission as "...to develop and deliver educational programs that equip limited-resource families and youth who are at risk for not meeting basic human needs, to lead positive, productive, contributing lives." CYFER offers resources for, and statistics about, child-youth-family development and programs, including full-text versions of pertinent journals and newsletters. Gopher to cyfer.esusda.gov.

ERIC Clearinghouse on Assessment and Evaluation - The ERIC_AE Gopher features a wide range of current essays about assessment and evaluation, a special Test Locator Service, and pointers to places on the Internet where you can search ERIC databases. Of special interest is the Test Locator (see Doolittle, Halpern, and Rudner, 1994). ERIC/AE, the Educational Testing Service (ETS), the Buros Institute, and Pro-Ed (publishing) have collaborated to produce several searchable testing databases. Featured is the ETS Test Collection containing descriptions of over 10,000 educational and psychological measures. Gopher to gopher.cua.edu and select Special Resources/ERIC Clearinghouse on Assessment.

The Library of Congress - LC offers the Machine-Assisted Realization of the Virtual Electronic library. Choose: Search LC Marvel Menus/Search LC Marvel Menus using Jughead and enter the search terms of your choice (e.g., dyslexia, cancer). Gopher to marvel.loc.gov

The Cornucopia of Disability Information - CODI offers 22 menu items of services covering a broad range of information for people with disabilities and for those who work with them. Gopher to val-dor.cc.buffalo.edu

The ERIC Clearinghouse on Instructional Technology - AskERIC maintains a large Gopher site providing access to a wide range of material. Of particular interest are the AskERIC Infoguides. Each Infoguide includes pertinent ERIC document citations and various Internet resources, such as appropriate listservs and pointers to gopher/

ftp sites. Some relevant titles are School Counseling 1 & 2, AIDS Education, Child Abuse, Disabilities, Disabled Students, Hotlines_Helplines, Sex Education, Special Education, and Vocational Education. Gopher to ericir.syr.edu.

Because Gopher is so popular, most large computer centers make Gopher available to their dial-up users. Often you can access Gopher at the system prompt by typing Gopher and the gopher address. For example, to access the AskERIC Gopher from a VAX computer you would type GOPHER ERICIR.SYR.EDU from the $.

Eric Databases

ERIC Abstracts Database - The entire contents of the Resources in Education and Current Index to Journals in Education are available though several Internet sites. You can Gopher to suvm.syr.edu and select database/ERIC/or gopher to gopher.uic.edu and select Library/Databases/ERIC.

ERIC Digest File - Digests are 1500 word reports that synthesize research and ideas about emerging issues in education. They are designed to help members of the educational community keep up-to-date with trends and new developments. Digests typically either serve as an introduction to a topic, provide current information of a factual nature related to a topic, define and describe a controversial topic, provide specific, concrete examples of how practitioners can apply research results in practical settings, report on the current status of research in an area, or summarize an existing review and synthesis publication. These are one of the most popular items on the entire Internet. Gopher to gopher.cua.edu and select ERIC Clearinghouse on Assessment/Search ERIC/, or gopher to gopher.ed.gov and select OERI/ERIC/Search Digests/.

AskERIC E-Mail Service

AskERIC is a personalized Internet-based service for educators and professionals allied with education support services. E-mail them a question and within 48 hours they will provide you with a response. Responses include full-text resources, ERIC database searches, learned opinions, and pointers to other resources. Send your e-mail to askeric@ericir.syr.edu

References

Doolittle, P., M. Halpern & L.M. Rudner (1994). The ERIC/AE Test Locator Service. *Educational Researcher*, 22, 8, pp. 34-35.

The following Internet references are available electronically by gophering to dewey.lib.ncsu.edu, select NCSU's "Library without Walls"/Reference Desk/Guides.

Electronic Frontier Foundation (1993). Big Dummy's Guide to the Internet, [Online].Available Gopher:vega.lib.ncsu.ed Directory / library/reference/guides, File: Guide.txt

Kehoe, B.P (1992). Zen and the Art of the Internet: A Beginners Guide to the Internet, (2nd Ed.), [Online]. Available FTP (Telnet): quake.think.com Directory:pub/etext/1992, File: Zen10.text

Polly, J.A. (1992). Surfing the Internet: An Introduction, (2nd.Ed.) [Online].Available Gopher:vega.lib.ncsu.edDirectory: library/ reference/guides, File: Surfing.txt

Liselle Drake is a User Services Coordinator and Lawrence Rudner is the Director of the ERIC Clearinghouse on Assessment and Evaluation at The Catholic University of America in Washington, DC.

ERIC Digest # EDO-CG-95-31

Locating and Evaluating Career Assessment Instruments

Jerome T. Kapes

For the purpose of both locating and evaluating career assessment instruments, there are three primary sources. Best known among these are the Buros Institute's publication *Tests in Print* and its comparison set of reviews in the *Mental Measurements Yearbooks* (MMY). A second source, which first became available in 1983-84, is *Tests and Test Critiques* (TC). Both include a listing and brief description of most tests commercially available in English-speaking countries (i.e., *Tests in Print* and *Tests*) as well as periodically published volumes of test reviews (i.e., MMY & TC)

The third source, which is published by the National Career Development Association, is *A Counselor's Guide to Career Assessment Instruments*. This book, first published in 1982 and every six years since, contains reviews of the most prominent career assessment instruments as well as brief descriptions of most others commercially available. In addition, this book also includes chapters on selecting, evaluating, using, and interpreting career relevant tests.

There are a number of other sources that focus on specialized aspects of career assessment. Also, certain journals publish test reviews or articles that provide evidence of the quality of specific career assessment instruments. These qualities typically are evaluated under the categories of norms, reliability, and validity. The American Educational Research Association (AERA), American Psychological Association (APA) and National Council on Measurement in Education (NCME) jointly publish *Standards for Educational and Psychological Testing*, which provides guidance for both test publishers and users of all types of tests (AERA, APA & NCME, 1985). In addition, the American Counseling Association (formerly the American Association for Counseling and Development, 1989) has produced its own guidelines, *Responsibilities of Users of Standardized Tests*, which provides additional help for users in counseling situations.

Much information is available to help locate and evaluate career assessment instruments, but it is the users who must employ these resources to make their own judgments about the appropriateness of a particular instrument for a specific situation. The purpose of this digest is to help users locate and organize information that will improve their evaluations.

Locating Instruments

Although *Tests in Print IV* (1994) and *Tests* (1991) are the most comprehensive listings of all tests, those wishing to locate a career assessment instrument may find it more useful to consult *A Counselor's Guide to Career Assessment Instruments 3rd edition* (Kapes, Mastie and Whitfield, 1994). This recent edition contains reviews of 52 prominent instruments along with an Additional Instruments chapter which briefly describes an additional 245 instruments. The entire 297 instruments are listed alphabetically in a User's Matrix that categorizes each entry by Characteristics (achievement, aptitude, interest, values/ satisfaction/environments, career development/maturity, personality) and Use level (elementary, junior high/middle school, senior high, 2 or 4 year college, adult education/ training, business and industry/ employment, disabled or disadvantaged). Those interested in locating a test for a specific purpose can use this matrix to identify instruments that may be appropriate.

If the instruments selected in the initial search are among the 52 with a complete review, the user can consult the reviews to further narrow the choice. Each entry includes a section of publisher-provided information that includes target population, statement of purpose, titles of subtests, scales and scores, forms and levels, date of most recent edition, languages available, time, norm groups, results reported, format, scoring, computer software, costs, comments, and published reviews. The review section is divided into the following headings: Description, Use in Counseling, Technical Considerations, Computer-Based Version (if available), Overall Critique, and References. If the instruments on which further information is needed are not among the 52 reviewed, the Additional Instruments chapter can be consulted to obtain the publisher, date and intended population on any of the 245 additional instruments. In addition, citations for all reviews published in the *Mental Measurements Yearbooks*, *Test Critiques*, previous editions of *A Counselor's Guide* as well as several other sources are listed. A brief description for each test is also included.

For those instruments not reviewed in *A Counselor's Guide* (3rd edition) or for a second opinion, the reader can consult one or more reviews cited in either the Additional Instruments chapter or at the end of the publisher information. If a *Mental Measurements Yearbook* (MMY) is to be consulted, it is necessary to know both the volume (sixth, 1965 through eleventh, 1992) and test number. The MMY entries typically contain brief publisher information along with two independent reviews. The reviews themselves are not divided into sections and tend to focus primarily on psychometric characteristics. Additional references are also provided.

To access *Test Critiques* (TC), it is also necessary to know the volume (Volume 1, 1984 through Volume 10, 1994). Each of these entries is written by a single reviewer and is divided into five categories (Introduction, Practical Applications/Use, Technical Aspects, Critique, and References). In addition to MMY and TC it may also be useful to consult the first (1982) and second (1988) editions of *A Counselor's Guide* or any of the other sources listed in the chapter on Sources of Information about Testing and Career Assessment in the third edition, which is an annotated bibliography of sources.

Evaluating Instruments

Many sources exist that could aid a user to evaluate the potential usefulness of a career assessment instrument. The previously mentioned AERA, APA, and NCME (1985) document provides guidance for both test publishers and users in the form of essential, conditional, and secondary standards. The standards, for example, call for a technical manual to be made available by the publisher so that any user can obtain information about the norms, reliability, and validity of the instrument as well as other relevant topics. It should be pointed out that, although the publisher typically provides evidence of this type from studies conducted with subjects for whom the instrument is intended, it may be necessary for the user to obtain data from other sources that better reflect the use intended for a particular application. This can be done from studies published in the literature (e.g., in the *Measurement and Evaluation in Counseling and Development* journal or the *Career Development Quarterly*) or from studies conducted on data collected by the user.

In addition to the technical qualities of norms, reliability, and validity, there are many other qualities of a career assessment instrument that need to be evaluated before selection for a particular use. In his chapter in *A Counselor's Guide* on "Selecting a Career Assessment Instrument," Mehrens covers many of these, including types of scores and interpretation materials, appropriateness for various groups, and a host of practical issues (e.g., qualifications of users, time, costs, and publisher support). In addition, he provides a "Test Evaluation Outline" that is included here to assist the user to systematically identify and collect all information necessary to conduct an adequate evaluation (Mehrens, 1994, p. 29):

1. State your purpose for testing
2. Describe the group that will be tested (for example, age or grade)
3. Name of test
4. Author(s)

5. Publisher
6. Copyright date(s)
7. Purpose and recommended use as stated in the manual
8. Grade/age levels for which the instrument was constructed
9. Forms: Are equivalent forms available? What evidence is presented on the equivalence of the forms?
10. Format: comment on legibility, attractiveness, and convenience
11. Cost
12. Content of test and types of items used
13. Administration and timing requirements
14. Scoring processes available (e.g., machine scoring)
15. Types of derived scores available
16. Types and quality of norms
17. Adequacy of reliability evidence presented in the manual
18. Validity evidence
19. General quality of administrative, interpretative and technical manuals
20. Comments about the instrument by outside reviewers

Summary

There are many sources to aid in locating and evaluating career assessment instruments. The primary sources are *A Counselor's Guide to Career Assessment Instruments*, the *Mental Measurement Yearbooks*, and *Test Critiques*. The AERA, APA and NCME Standards provide guidance to publishers and users on the qualities of norms, reliability, and validity as well as many other considerations that affect test use. However, the bottom line is that the user is responsible for making the final judgment about the appropriateness of a particular instrument for a specific use.

References

American Association for Counseling and Development (1989). *Responsibilities of users of standardized tests* (AACD/AMECD Policy Statement). Alexandria, VA: Author.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Washington, D.C.: American Psychological Association.

Kapes, J.T., Mastie, M.M., & Whitfield, E.A. (Eds.). (1994). *A counselor's guide to career assessment instruments* (3rd edition). Alexandria, VA: National Career Development Association.

Keyser, J.D., & Sweetland, R.C. (1984-94). Test critiques (Vols. 1-10). Austin, TX: PRO-ED.

Kramer, J.J., & Conoley, J.C. (Eds.). (1992). *Eleventh mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Mehrens, W.A. (1994). Selecting a career assessment instrument. In J.T. Kapes, M.M. Mastie, & E.A. Whitfield (Eds.), *A counselor's guide to career assessment instruments* (3rd edition). Alexandria, VA: National Career Development Association.

Murphy, L.L., Conoley, J.C., Impara, J.C., & Mitchell, J.V. (Eds.). (1994). *Tests in print IV*. Lincoln, NE: Buros Institute of Mental Measurement.

Sweetland, R.C., & Keyser, D.J. (Eds.). (1991). *Tests: A comprehensive reference for assessments in psychology, education, and business* (3rd edition). Austin, TX: PRO-ED.

Jerome T. Kapes, Ph.D., NCC, NCCC, is a professor of Educational Psychology and a faculty member in the Counseling Psychology, Educational Psychology Foundations and Vocational Education programs at Texas A&M University in College Station, Texas.

ERIC Digest # EDO-CG-95-32

Inappropriate Statistical Practices in Counseling Research: Three Pointers for Readers of Research Literature

Bruce Thompson

The research literature provides important guidance to counselors working to keep abreast of the latest thinking regarding best practices and recently developed counseling tools. However, in my work as a former editor of *Measurement and Evaluation in Counseling and Development*, and as editor of *Journal of Experimental Education* and of *Educational and Psychological Measurement*, I have noticed some errors that seem to recur within the research literature read by counselors. The purpose of this digest is to highlight a few of these errors, and to provide some helpful references that further explore these problems. In "buying" the ideas presented within publications, as in buying more tangible products, the old maxim of *caveat emptor* does indeed remain useful.

1. Insufficient Attention to Score Reliability

Pedhazur and Schmelkin (1991, pp. 2-3) recently noted that, "Measurement is the Achilles' heel of sociobehavioral research... [I]t is, therefore, not surprising that little or no attention is given to properties of measures used in many research studies." In fact, empirical studies of the published literature indicate that score reliability is not considered in between 40 and 50 percent of the published research. And, similarly, in doctoral dissertations we occasionally even see scores being analyzed that have reliability coefficients that are *less* than *negative* one (Thompson, 1994)!

The failure to consider score reliability adequately in substantive research is very serious, because effect sizes and power against Type II errors are both attenuated by measurement error. Thus, prospectively we may plan and conduct studies that could not possibly yield noteworthy effectsizes, given that score unreliability inherently attenuates effect sizes. Or, retrospectively, we may not accurately interpret the effect sizes in completed studies if we do not consider as part of our interpretation the reliability of the scores we are actually analyzing.

Consumers of published research should generally expect authors to analyze the reliability of the scores in their own data. It is not sufficient even to report reliability coefficients from test manuals or from other research, because tests are not themselves reliable (i.e., tests are not imprinted both with ink and with reliability during the various stages of the printing process). Score reliability is influenced by various facets of the measurement process, including when, how, and to whom the test was administered. Thus, it becomes an oxymoron to speak of "the reliability of the test," because such a telegraphic shorthand way of speaking is also an incorrect way of speaking, i.e., makes an inherently untrue assertion.

Partly because this shorthand way of speaking is so common, too few researchers recognize that *reliability is a characteristic of scores* and not of tests. Because scores possess or lack these characteristics, different sets of scores generated by even the same measure may each have different reliabilities.

These telegraphic ways of speaking become problematic if we come unconsciously to ascribe literal truth to our shorthand, rather than recognizing that our jargon is sometimes literally untrue. As noted elsewhere:

> This is not just an issue of sloppy speaking—the problem is that sometimes we unconsciously come to think what we say or what we hear, so that sloppy speaking does sometimes lead to a more pernicious outcome, sloppy thinking and sloppy practice (Thompson, 1992, p. 436).

Readers of published research should expect authors only to offer assertions that they reasonably believe are true, and thus we should not condone use of the language, "the test is reliable." Furthermore, we should expect authors of published research to offer empirical evidence that the scores they are actually analyzing have reasonable measurement integrity.

2. Overreliance on Tests of Statistical Significance

The business of science is identifying relationships that recur under stated conditions. Unhappily, too many researchers at least unconsciously incorrectly assume that the p values calculated in statistical significance tests evaluate the probability that results will recur (Carver, 1993).

To get a single estimate of the p(robability) of the sample statistics, the null hypothesis is posited to be exactly true in the population. Thus, statistical significance testing evaluates *the probability of the sample statistics for the data in hand, given that the null hypothesis about the related parameters in the population is presumed to be exactly true*. This is *not* a test of result replicability, i.e., is not a test of whether roughly equivalent effect sizes would be detected in subsequent studies conducted under similar conditions!

In fact, the requirement that statistical significance testing must presume an assumption that the null hypothesis is true in the population is a requirement that an untruth be presumed. As Meehl (1978, p. 822) notes, "As I believe is generally recognized by statisticians today and by thoughtful social scientists, the null hypothesis, taken literally, is always false." Similarly, Hays (1981, p. 293) points out that "[t]here is surely nothing on earth that is completely independent of anything else [in the population]. The strength of association may approach zero, but it should seldom or never be exactly zero."

And positing an untruth about the population has a very important implication. Whenever the null is not *exactly* true in the sample(s), then the nullhypothesis will *always* be rejected at some sample size. As Hays (1981, p. 293) emphasizes, "virtually any study can be made to show significant results if one uses enough subjects."

Although statistical significance is a function of several different design features, sample size is a basic influence on statistical significance. Thus, statistical significance testing can create a tautology in which we invest energy to determine that which we already know, i.e., our sample size.

Consumers of published research should expect authors to never say "significant" when they mean "statistically significant." Since statistical significance does not evaluate result importance, always using the phrase "statistically significant" when referring to statistical tests helps somewhat to avoid confusing statistical significance with the issue of importance. As Thompson (1993) emphasized:

> Statistics can be employed to evaluate the probability of an event. But importance is a question of human values, and math cannot be employed as an atavistic escape (a la Fromme's Escape from Freedom) from the existential human responsibility for making value judgments. If the computer package did not ask you your values prior to its analysis, it could not have considered your value system in calculating *p*'s, and so *p*'s cannot be blithely used to infer the value of research results. Like it or not, empirical science is inescapably a subjective business. (p. 365)

Second, it is important to expect authors reporting statistical significance to supplement these tests with analyses that do focus on result importance and on result replicability. With respect to result importance, authors should be expected to report and interpret effect sizes. Even the recently published fourth edition APA style manual acknowledges that probability values reflect sample size, and thus encourages all authors to provide effect-size information.

With respect to result replicability, authors should be expected to report actual, so-called "external" replication studies, or to conduct "internal" replicability analyses (Thompson, 1993, 1994b). The later include cross-validation, the jackknife, and the bootstrap. These analyses, unlike statistical significance tests, do inform judgments about whether detected relationships replicate under stated conditions.

3. Stepwise Methods Should Not Be Used

Stepwise analyses are used with some frequency in published research, almost always to bad effect (cf. Thompson, 1994a). There are three problems. First, the computer packages use the wrong degrees of freedom in computing statistical significance in these analyses, and the incorrect degrees of freedom systematically bias the tests in favor of yielding statistical significance that is bogus. Second, not only does doing k steps of analysis not yield the best predictor set of size *k*, it can occur that none of the predictors entered in the first *k* steps are even among the best predictor set of size *k*. Third, because the linear sequence of entry decisions can be radically influenced by sampling error, thus throwing the whole sequence of decisions off track at any step, and because so many decisions are made along the way of a stepwise analysis, stepwise analyses often produce results that are very unlikely to replicate!

References

Carver, R. (1993). The case against statistical significance testing, revisited. *Journal of Experimental Education, 61*(4), 287-292.

Hays, W. L. (1981). Statistics (3rd ed.). New York: Holt, Rinehart and Winston.

Meehl, P.E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46*, 806-834.

Pedhazur, E.J., & Schmelkin, L.P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Erlbaum.

Thompson, B. (1992). Two and one-half decades of leadership in measurement and evaluation. *Journal of Counseling and Development, 70*, 434-438.

Thompson, B. (1993). The use of statistical significance tests in research: Bootstrap and other alternatives. *Journal of Experimental Education, 61*(4), 361-377.

Thompson, B. (1994, April). Common methodology mistakes in dissertations, revisited. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED forthcoming)

Thompson, B. (1994b). The pivotal role of replication in psychological research: Empirically evaluating the replicability of sample results. *Journal of Personality*, *62*(2), 157-176.

Bruce Thompson is Professor of Education and Distinguished Research Fellow at Texas A&M University, and Adjunct Professor of Community Medicine at Baylor College of Medicine in Houston.

ERIC Digest # EDO-CG-95-33

ASSESSMENT
Issues & Challenges for the Millennium

This book offers an intriguing collection of papers that provide an update on new developments in assessment as well as discussions of important topics for counselor/educators, practitioners, scholars and researchers.

The papers are grouped into five sections: Emerging Issues, Counselor Education, K-12 Guidance and Counseling, Special Populations and Special Topics. In addition, an appendix of 33 digests on Assessment in Counseling and Therapy.

Assessment - Issues & Challenges for the Millennium is a compilation of incredibly practical pieces specifically addressing some of the most pressing assessment topics faced by today's professional counselors. The book is a "must read" for all mental health professionals and was written by some of the most recognized assessment experts within the profession.

Gerald A. Juhnke, Ph.D.
Professor, University of North Carolina at Greensboro
President, Association for Assessment in Counseling

This book provides one solid resource and a great opportunity for counselors to hear the views of assessment practitioners on current issues.

Janet Wall, Ed.D.
Assistant Vice President of ACT, Head of Educational Technology
President-Elect, Association for Assessment in Counseling

ISBN 1-56109-091-3
9 781561 090914

ERIC Counseling & Student Services Clearinghouse
201 Ferguson Building • School of Education
University of North Carolina at Greensboro
Greensboro, North Carolina
27402 • 800-414-9769
ericcass@uncg.edu
FAX: 336/334-4116